



**The Journal of Robotics,
Artificial Intelligence & Law**

Editor's Note: The Future
Victoria Prussen Spears

The Future of AI Protection
Paul A. Ragusa and Nick Palmieri

Should the Use of Lethal Autonomous Robots Be Permitted in Warfare?
Jake Okechukwu Effoduh

NIST Solicits Comments on Four Principles of Explainable Artificial Intelligence and Other Developments

Lee J. Tiedrich, Sam Jungyun Choi, and James Yoon

NHTSA Continues Its Focus on Advancing Autonomous Vehicle Technologies
Rebecca Baden Chaney, Cheryl A. Falvey, and Rukiya Mohamed

Horse Cases, the Cheapest Cost Avoider Rule, and Liability for Highly Autonomous Vehicle Accidents
Daniel Barabander

Everything Is Not *Terminator*: Assessment of Artificial Intelligence Systems
John Frank Weaver

- 5 Editor’s Note: The Future**
Victoria Prussen Spears
- 7 The Future of AI Protection**
Paul A. Ragusa and Nick Palmieri
- 17 Should the Use of Lethal Autonomous Robots Be Permitted in Warfare?**
Jake Okechukwu Effoduh
- 29 NIST Solicits Comments on Four Principles of Explainable Artificial Intelligence and Other Developments**
Lee J. Tiedrich, Sam Jungyun Choi, and James Yoon
- 33 NHTSA Continues Its Focus on Advancing Autonomous Vehicle Technologies**
Rebecca Baden Chaney, Cheryl A. Falvey, and Rukiya Mohamed
- 43 Horse Cases, the Cheapest Cost Avoider Rule, and Liability for Highly Autonomous Vehicle Accidents**
Daniel Barabander
- 67 Everything Is Not *Terminator*: Assessment of Artificial Intelligence Systems**
John Frank Weaver

EDITOR-IN-CHIEF

Steven A. Meyerowitz

President, Meyerowitz Communications Inc.

EDITOR

Victoria Prussen Spears

Senior Vice President, Meyerowitz Communications Inc.

BOARD OF EDITORS

Miranda Cole

Partner, Covington & Burling LLP

Kathryn DeBord

Partner & Chief Innovation Officer, Bryan Cave LLP

Melody Drummond Hansen

Partner, O'Melveny & Myers LLP

Paul B. Keller

Partner, Allen & Overy LLP

Garry G. Mathiason

Shareholder, Littler Mendelson P.C.

Elaine D. Solomon

Partner, Blank Rome LLP

Linda J. Thayer

Partner, Finnegan, Henderson, Farabow, Garrett & Dunner LLP

Edward J. Walters

Chief Executive Officer, Fastcase Inc.

John Frank Weaver

Attorney, McLane Middleton, Professional Association

THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW (ISSN 2575-5633 (print) /ISSN 2575-5617 (online) at \$495.00 annually is published six times per year by Full Court Press, a Fastcase, Inc., imprint. Copyright 2021 Fastcase, Inc. No part of this journal may be reproduced in any form—by microfilm, xerography, or otherwise—or incorporated into any information retrieval system without the written permission of the copyright owner. For customer support, please contact Fastcase, Inc., 711 D St. NW, Suite 200, Washington, D.C. 20004, 202.999.4777 (phone), 202.521.3462 (fax), or email customer service at support@fastcase.com.

Publishing Staff

Publisher: Morgan Morrisette Wright

Journal Designer: Sharon D. Ray

Cover Art Design: Juan Bustamante

Cite this publication as:

The Journal of Robotics, Artificial Intelligence & Law (Fastcase)

This publication is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional services. If legal advice or other expert assistance is required, the services of a competent professional should be sought.

Copyright © 2021 Full Court Press, an imprint of Fastcase, Inc.

All Rights Reserved.

A Full Court Press, Fastcase, Inc., Publication

Editorial Office

711 D St. NW, Suite 200, Washington, D.C. 20004

<https://www.fastcase.com/>

POSTMASTER: Send address changes to THE JOURNAL OF ROBOTICS, ARTIFICIAL INTELLIGENCE & LAW, 711 D St. NW, Suite 200, Washington, D.C. 20004.

Articles and Submissions

Direct editorial inquiries and send material for publication to:

Steven A. Meyerowitz, Editor-in-Chief, Meyerowitz Communications Inc.,
26910 Grand Central Parkway, #18R, Floral Park, NY 11005, smeyerowitz@
meyerowitzcommunications.com, 646.539.8300.

Material for publication is welcomed—articles, decisions, or other items of interest to attorneys and law firms, in-house counsel, corporate compliance officers, government agencies and their counsel, senior business executives, scientists, engineers, and anyone interested in the law governing artificial intelligence and robotics. This publication is designed to be accurate and authoritative, but neither the publisher nor the authors are rendering legal, accounting, or other professional services in this publication. If legal or other expert advice is desired, retain the services of an appropriate professional. The articles and columns reflect only the present considerations and views of the authors and do not necessarily reflect those of the firms or organizations with which they are affiliated, any of the former or present clients of the authors or their firms or organizations, or the editors or publisher.

QUESTIONS ABOUT THIS PUBLICATION?

For questions about the Editorial Content appearing in these volumes or reprint permission, please contact:

Morgan Morrisette Wright, Publisher, Full Court Press at mwright@fastcase.com
or at 202.999.4878

For questions or Sales and Customer Service:

Customer Service

Available 8 a.m.–8 p.m. Eastern Time

866.773.2782 (phone)

support@fastcase.com (email)

Sales

202.999.4777 (phone)

sales@fastcase.com (email)

ISSN 2575-5633 (print)

ISSN 2575-5617 (online)

NIST Solicits Comments on Four Principles of Explainable Artificial Intelligence and Other Developments

Lee J. Tiedrich, Sam Jungyun Choi, and James Yoon*

The authors of this article discuss a National Institute of Standards and Technology white paper that seeks to define the principles that capture the fundamental properties of explainable artificial intelligence systems and related legislative developments.

The National Institute of Standards and Technology (“NIST”) released the first draft of the Four Principles of Explainable Artificial Intelligence,¹ a white paper that seeks to define the principles that capture the fundamental properties of explainable artificial intelligence (“AI”) systems. NIST accepted comments through October 15, 2020.

The White Paper

In February 2019, the Executive Order on Maintaining American Leadership in Artificial Intelligence² directed NIST to develop a plan that would, among other objectives, “ensure that technical standards minimize vulnerability to attacks from malicious actors and reflect Federal priorities for innovation, public trust, and public confidence in systems that use AI technologies; and develop international standards to promote and protect those priorities.”

In response, NIST issued a plan³ in August 2019 for prioritizing federal agency engagement in the development of AI standards, identifying seven properties that characterize trustworthy AI—accuracy, explainability, resiliency, safety, reliability, objectivity, and security.

NIST’s white paper focuses on explainability and identifies four principles underlying explainable AI.

- *Explanation.* AI systems must supply evidence, support, or reasoning for their outputs. Researchers have developed different models to explain AI systems, such as self-explainable models where the models themselves are the provided explanation.
- *Meaningful.* The recipient must understand the AI system's explanation. This principle is a contextual requirement—for example, different types of user groups may require different explanations, or a particular user's prior knowledge, experiences, and mental processes may affect meaningfulness. Hence, tailoring is necessary for effective communication.
- *Explanation Accuracy.* The explanation must correctly reflect the AI system's process for generating its output. In contrast to decision accuracy, explanation accuracy is not concerned with whether or not the system's judgment is correct. It is referencing how the system came to its conclusion. The principle is also contextual—there may be different explanation accuracy metrics for different types of groups and users.
- *Knowledge Limits.* The AI system must identify cases it was not designed or approved to operate, or where its answers are not reliable. This ensures that reliance on AI system's decision processes occurs only where it is appropriate.

The white paper states that explanations generally can be described along two dimensions: the amount of time the consumer has to respond to the information and the level of detail in the explanation. Although flexibility in the range and types of explanations will be necessary, NIST provides a non-exhaustive list of explanation categories, drawing from academic literature:

- *User Benefit.* This type of explanation is designed to inform a user about an AI system output, such as providing the reason a loan application was approved or denied to the applicant.
- *Societal Acceptance.* This type of explanation is designed to generate trust and acceptance by society, to provide an increased sense of comfort in the system.
- *Regulatory and Compliance.* This type of explanation assists with audits for compliance with regulations, standards, and

legal requirements, such as providing detailed explanation to a safety regulator to evaluate the output of self-driving cars.

- *System Development*. This type of explanation assists with developing, improving, debugging, or maintaining an AI system by technical staff and product managers.
- *Owner Benefit*. This type of explanation benefits the operator of a system, such as a recommendation system that lists movies to watch and explains the selection based on previously viewed items.

After explaining the core concepts of explainable AI systems, NIST explores the explainability of human decision processes. NIST states that humans demonstrate only a limited ability to meet the four principles described above, which provides a benchmark to evaluate explainable AI systems and informs the development of reasonable metrics. According to NIST, evaluating explainability in context of human decision-making also may lead to better understanding of human-machine collaboration and interfaces.

Although the white paper does not provide detailed guidance for organizations implementing AI systems, it represents an important step by NIST to develop trustworthy AI tools. Documents from other jurisdictions on explaining AI provide more detailed guidance aimed at helping organizations operationalize the concept of explainable AI.

The UK Information Commissioner's Office ("ICO"), for example, issued on May 20, 2020, its final guidance on explaining decisions made with AI.⁴ Similar to the NIST white paper, the ICO recognizes that there are different underlying principles to be followed and different models of AI explanation. The ICO takes these principles one step further, however, and provides more detailed guidance on how to explain AI in practice, depending on the type of AI system used.

Some Legislative Developments Relating to NIST

Efforts to advance the development of AI standards through NIST has been a topic of increasing focus in Congress. Recent bills include Senator Cory Gardner's (R-CO) Advancing Artificial Intelligence Research Act of 2020,⁵ which would appropriate

\$250 million to NIST for each of fiscal years 2021 through 2025 for the creation of a national program to advance AI research, and Representative Eddie Bernice Johnson's (D-TX-30) National Artificial Intelligence Initiative Act of 2020,⁶ which would appropriate over \$50 million to NIST for each of fiscal years 2021 through 2025 for the research and development of voluntary standards for trustworthy AI systems, among other activities.

The House Appropriations Committee also released the draft fiscal year 2021 Commerce, Justice, Science, and Related Agencies funding bill,⁷ where \$789 million is included for core NIST research activities, an increase of \$35 million above the FY 2020 enacted level.

Notes

* Lee J. Tiedrich (ltiedrich@cov.com) is an IP/technology transactions partner at Covington & Burling LLP and co-chair of the firm's global and multi-disciplinary Artificial Intelligence Initiative. Sam Jungyun Choi (jchoi@cov.com) is an associate in the firm's technology regulatory group. James Yoon (jyoon@cov.com) is an associate in the firm's Data Privacy and Cybersecurity Practice Group.

1. NISTIR 8312, [https://www.nist.gov/system/files/documents/2020/08/17/NIST Explainable AI Draft NISTIR8312 %281%29.pdf](https://www.nist.gov/system/files/documents/2020/08/17/NIST%20Explainable%20AI%20Draft%20NISTIR8312%20%281%29.pdf).

2. <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>.

3. https://www.nist.gov/system/files/documents/2019/08/10/ai_standards_fedengagement_plan_9aug2019.pdf.

4. <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence-1-0.pdf>.

5. <https://www.congress.gov/bill/116th-congress/senate-bill/3891?q=%7B%22search%22%3A%5B%22chamberActionDateCode%3A%5C%222020-06-04%7C116%7C10000%5C%22+AND+billIsReserved%3A%5C%22N%5C%22%22%5D%7D&r=3&s=1>.

6. <https://www.congress.gov/bill/116th-congress/house-bill/6216>.

7. <https://appropriations.house.gov/sites/democrats.appropriations.house.gov/files/CJS.pdf>.