



IEEE FEDERATED MACHINE LEARNING WHITE PAPER

Authored by

Qiang Yang

Lixin Fan

Richard Tong

Angelica Lv

TRADEMARKS AND DISCLAIMERS

IEEE believes the information in this publication is accurate as of its publication date; such information is subject to change without notice. IEEE is not responsible for any inadvertent errors.

The ideas and proposals in this specification are the respective author's views and do not represent the views of the affiliated organization.

The Institute of Electrical and Electronics Engineers, Inc. 3 Park Avenue, New York, NY 10016-5997, USA

Copyright © 2021 by The Institute of Electrical and Electronics Engineers, Inc.

All rights reserved. June 2021. Printed in the United States of America.

PDF: STDVA24772 978-1-5044-7660-7

IEEE is a registered trademark in the U. S. Patent & Trademark Office, owned by The Institute of Electrical and Electronics Engineers, Incorporated. All other trademarks are the property of the respective trademark owners.

IEEE prohibits discrimination, harassment, and bullying. For more information, visit <http://www.ieee.org/web/aboutus/whatis/policies/p9-26.html>.

No part of this publication may be reproduced in any form, in an electronic retrieval system, or otherwise, without the prior written permission of the publisher.

Find IEEE standards and standards-related product listings at: <http://standards.ieee.org>.

NOTICE AND DISCLAIMER OF LIABILITY CONCERNING THE USE OF IEEE SA DOCUMENTS

This IEEE Standards Association (“IEEE SA”) publication (“Work”) is not a consensus standard document. Specifically, this document is NOT AN IEEE STANDARD. Information contained in this Work has been created by, or obtained from, sources believed to be reliable, and reviewed by members of the activity that produced this Work. IEEE and the IEEE P3652.1 members expressly disclaim all warranties (express, implied, and statutory) related to this Work, including, but not limited to, the warranties of: merchantability; fitness for a particular purpose; non-infringement; quality, accuracy, effectiveness, currency, or completeness of the Work or content within the Work. In addition, IEEE and the IEEE P3652.1 members disclaim any and all conditions relating to: results; and workmanlike effort. This document is supplied “AS IS” and “WITH ALL FAULTS.”

Although the IEEE P3652.1 members who have created this Work believe that the information and guidance given in this Work serve as an enhancement to users, all persons must rely upon their own skill and judgment when making use of it. IN NO EVENT SHALL IEEE-SA OR ICAP MEMBERS BE LIABLE FOR ANY ERRORS OR OMISSIONS OR DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO: PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS WORK, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE AND REGARDLESS OF WHETHER SUCH DAMAGE WAS FORESEEABLE.

Further, information contained in this Work may be protected by intellectual property rights held by third parties or organizations, and the use of this information may require the user to negotiate with any such rights holders in order to legally acquire the rights to do so, and such rights holders may refuse to grant such rights. Attention is also called to the possibility that implementation of any or all of this Work may require use of subject matter covered by patent rights. By publication of this Work, no position is taken by the IEEE with respect to the existence or validity of any patent rights in connection therewith. The IEEE is not responsible for identifying patent rights for which a license may be required, or for conducting inquiries into the legal validity or scope of patents claims. Users are expressly advised that determination of the validity of any patent rights, and the risk of infringement of such rights, is entirely their own responsibility. No commitment to grant licenses under patent rights on a reasonable or non-discriminatory basis has been sought or received from any rights holder.

This Work is published with the understanding that IEEE and the IEEE P3652.1 members are supplying information through this Work, not attempting to render engineering or other professional services. If such services are required, the assistance of an appropriate professional should be sought. IEEE is not responsible for the statements and opinions advanced in this Work.

TABLE OF CONTENTS

IEEE FEDERATED MACHINE LEARNING WHITE PAPER.....	5
ABSTRACT	5
1. DATA PRIVACY POROTECTION AND FEDERATED MACHINE LEARNING	6
1.1. DATA ISOLATION AND PRIVACY PROTECTION.....	6
1.2. FEDERATED MACHINE LEARNING FOR DATA PRIVACY PROTECTION.....	6
1.3. WHY IEEE FEDERATED MACHINE LEARNING (IEEE STD 3652.1-2020)	7
2. FRAMEWORK AND CATEGORIZATION OF FEDERATED MACHINE LEARNING	7
3. APPLICATIONS OF FEDERATED MACHINE LEARNING	10
3.1. FINANCE.....	10
3.2. TELECOMMUNICATIONS	10
3.3. HEALTHCARE	11
3.4. EDUCATION.....	11
3.5. URBAN COMPUTING.....	13
3.6. GOVERNMENT SERVICES.....	14
3.7. GOVERNMENT GOVERNANCE	14
3.8. MARKETING	14
3.9. IOT/EDGE COMPUTING.....	15
4. REFERENCES.....	17

IEEE FEDERATED MACHINE LEARNING WHITE PAPER

ABSTRACT

Data privacy and information security pose significant challenges to the big data and artificial intelligence (AI) community as these communities are increasingly under pressure to adhere to regulatory requirements, such as the European Union's General Data Protection Regulation. Many routine operations in big data applications, such as merging user data from various sources in order to build a machine learning model, are considered to be illegal under current regulatory frameworks. The purpose of federated machine learning is to provide a feasible solution that enables machine learning applications to utilize the data in a distributed manner that does not exchange raw data directly and does not allow any party to infer private information of other parties. This white paper intends to present an overview of the Federated Machine Learning (FML) technology that can be used as a basis for standards, certifications, laws, policies, and/or product ratings.

This white paper targets an educated audience, including lawmakers, corporate and governmental policy makers, manufacturers, engineers, and standard setting bodies. However, this white paper is also easily understood by non-technical managers and policy makers as it provides system developers and manufacturers with an overview of Federated Machine Learning techniques. Finally, one must give credit to the IEEE Federated Machine Learning (P3652.1) working group participants for their tremendous dedication, expertise and thoughtful collaborations, without which the publication of IEEE Std 3652.1-2020 [1] would not have been possible .

1. DATA PRIVACY PROTECTION AND FEDERATED MACHINE LEARNING

1.1. DATA ISOLATION AND PRIVACY PROTECTION

Today, artificial intelligence (AI) technology is showing its strengths in almost every industry and walks of life. The current public interest in AI is partly driven by Big Data availability, e.g., AlphaGo in 2016 used a total of 300,000 games as training data and achieved excellent results. However, except for a few industries, most application fields have only limited data of poor quality, making the realization of AI technology more difficult than previously thought. Moreover, data privacy and information security pose significant challenges to the big data and AI community as these communities are increasingly under pressure to adhere to regulatory requirements such as the European Union's General Data Protection Regulation. Many routine operations in big data applications, such as merging user data from various sources to build a machine learning model, are deemed illegal under current regulatory frameworks.

As a result, we face a dilemma that our data is in the form of isolated islands, but we are forbidden in many situations to collect, fuse, and use the data from different places for AI processing. How to legally solve the problem of data fragmentation and isolation is a major challenge for AI researchers and practitioners today.

1.2. FEDERATED MACHINE LEARNING FOR DATA PRIVACY PROTECTION

The concept of federated learning was initially proposed by Google researchers to build machine learning models based on data sets that are distributed across multiple mobile phone devices while preventing data leakage [2], [3]. The idea was extended by researchers from China, Singapore, Europe, and the United States to cover secure distributed and collaborative learning scenarios among multiple organizations such as banks and hospitals that have respective private data but do not wish to have this data shared [4]. Organizations such as banks or health centers would like to take advantage of machine learning models co-developed with peer organizations but are obligated to keep their own data under tight protection.

Federated machine learning is a technological framework that allows a machine learning model to be collectively constructed and used through data that is distributed across repositories owned by different organizations or devices. While facilitating the building of federated machine learning models, this framework also aims to preserve privacy, improve security, and meet regulatory requirements concerning data usage.