



ONNX |

Workshop
2020



ONNX

Welcome!

Disclaimer

All workshop presentations, SIG/WG sessions will be recorded and made available publicly afterwards.

Here today ... 200+ registrants, 100+ organizations

Accenture	Couchbase	Leidos BioMedical	REdu
Adobe	Cruise	LF AI Foundation	rollApp Inc.
AI Startup	Datarobot	LF Energy	SafeGraph
Alamo Colleges	EP Ehler Prüftechnik Engineering	Liftr Insights	SAP
Alibaba Group	Ericsson	Living Open Source Zambia LTD	Siemens
Altran	Esperanto Technologies	MathWorks	Spark CG
Amazon	Facebook	Maxar	Tech Mahindra Americas Inc
AMD	Faurecia Digital Service Factory	Metalsa India Pvt ltd	The Linux Foundation
Apache Heron	Frederick National Lab	Microchip	TheTeamPhoenix
ArangDB	Fulton	Microsoft	Thundersoft
arsia labs	Futurewei Technologies, Inc.	Mohammed V University	TJ Solutions KFT
Arthur AI	Gartner	Mt Siani	Trace3
Aselsan	GE Healthcare	National Cancer Institute	TUHH Audi
Bank of Montreal	Grab	Neurosoft Sp. z O.o.	UIUC
Baylibre SAS	Hcl	Nexstar Media Group/Boards	University of California Santa Cruz
BCEL	Hewlett Packard Enterprise	NIH/NCI	University of Exeter
Beckhoff Automation	Huawei	Nuance Communication Inc	University of Illinois Urbana-
Blaize New Computing Technologies	IBM	NVIDIA Corporation	US Turf
Bogota	Infosys	Oracle	UST Global Inc.
Bootlabs technology	Infy	Preferred Networks	Warburg Pincus
bunq	Instituto Federal de Ciência e	Publicis Sapient	Workselection
Check Point	Intel Corporation	Punjab Information Technology	www.microsemi.com
CIT	Intelligrape	Qualitas Technologies	XenonStack
ClearBox AI Solutions	John Deere	RDC Partner	Xilinx
Click2Cloud	Juniper Networks	RedHat	YPF S.A.
Cognitiviti	Kichakato Kizito		

Logistics

- Host of Zoom Meeting will share the slides on screen and record all presentations.
- All participants will be muted except when presenting.
- Questions should be posted in the “chat” box and will be answered during the Q&A section after each block of presentations.
- Please “raise hand” (Zoom feature) if you would like to speak and engage in the discussion.

Goals for the Workshop

- Get the latest updates on ONNX - Releases, Governance and SIGs/WGs
- Learn from the community and how ONNX is being used
- Share feedback on what is working (and what isn't)
- Learn how to get more involved with ONNX Steering Committee, SIGs and Working Groups

Agenda - Part 1

	ONNX Community Virtual Meetup	
	April 9, 2020 - Agenda	
9:00 - 9:05AM	<i>Opening & Welcome</i>	Tom Truong (IBM)
9:05 - 9:30AM	ONNX Community & LFAI Update	ONNX Steering Committee Ibrahim Haddad (LF AI)
	<i>Partner / End User Stories</i>	
9:30 - 9:40AM	ONNX Adoption and Internal Use Cases	Steven Eliuk (IBM)
9:40 - 9:50AM	MindSpore - DL Framework for ONNX/MLIR	Zhipeng Huang (Huawei)
9:50 - 10:00AM	ONNX Runtime Optimizations Breakthrough	Emma Ning (Microsoft)
10:00 - 10:10AM	FINN: Pytorch-to-FPGA Flow for QNNs	Yaman Umuroglu (Xilinx)
10:10 - 10:20AM	Study of Genome Assemblies using ONNX	Kishwar Shafin (UCSC)
10:20 - 10:30AM	Azure OCR Cognitive Service acceleration using ONNX	Yung-Shin Lin (Microsoft)
10:30 - 10:40AM	ONNX Use Cases at Mathworks.com	Shounak Mitra (Mathworks)
10:40 - 10:50AM	<i>Q&A / Open Discussions</i>	All
10:50 - 11:00AM	<i>Break</i>	

Agenda - Part 2

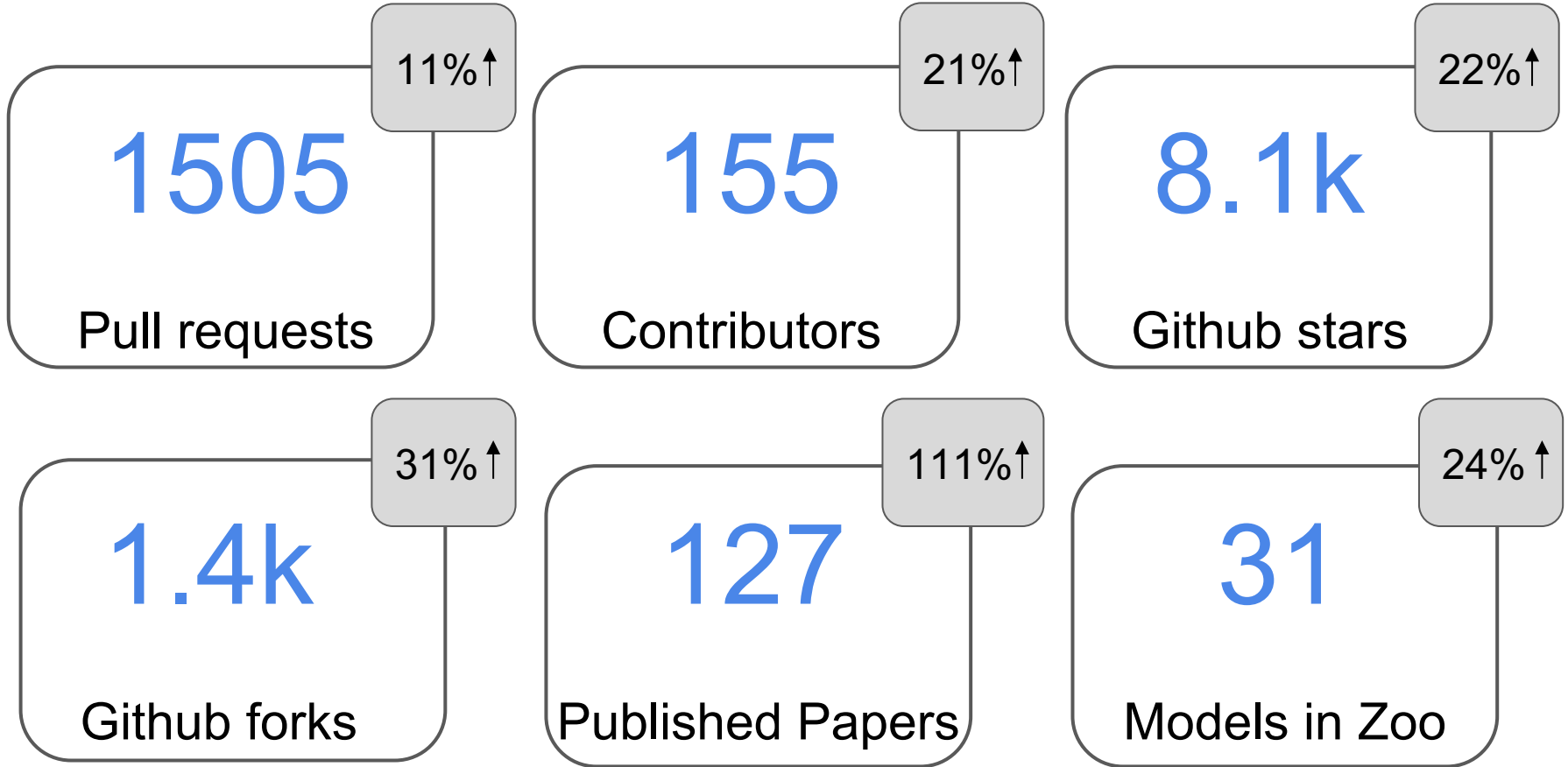
	SIG/WG Updates	
11:00 - 11:10AM	Architecture/Infrastructure SIG	Ganesan Ramalingam (Microsoft) Ke Zhang (Alibaba)
11:10 - 11:20AM	Operators	Michał Karzyński (Intel) Emad Barsoum (Microsoft)
11:20 - 11:30AM	Converters	Chin Huang (IBM) Guenther Schmuelling (Microsoft)
11:30 - 11:40AM	Model Zoo/Tutorials	Vinitra Swamy (Microsoft)
11:40 - 11:50AM	Training	Svetlana Levitan (IBM)
11:50 - 12:00PM	Q&A / Open Discussions	All
12:00 PM	<i>Wrap Up</i>	



ONNX

State of the state

Engagement & usage (compared to 3/22/19)



Support

Creation/ Manipulation



NEW



Run/ Compile



NEW



Visualization/ Test Tools



Coming soon: ONNX 1.7

ONNX v1.7 comes with exciting new and enhanced features!

- Model training, introduced as a tech preview, expands ONNX beyond original inference capabilities.
- A new protobuf message is added to describe training information, including the training algorithm, initializers, and new operators such as gradients, loss functions and optimizers.
- The new spec allows one to create a training model in one framework, export it in ONNX, and load into a runtime or another framework where the training can proceed.
- Operator registration APIs are updated to support dynamic function body (sub-graph) registration.
- Functions' body graph are extended to be able to rely on multiple external operator sets.
- The model checker is enhanced in shape inference and type constraints specified in the operator schema.
- Version numbers: IR = 7, opset = 12, opset-ml = 2, opset-training = 1

Thank you everyone for your countless hours of work!



ONNX

Governance

ONNX open governance recap

Steering Committee defines the vision, goals, and governance process of the ONNX community. Steering Committee is elected.

Special Interest Groups (SIGs) are **persistent groups** that are responsible for specific parts of the project, including ongoing maintenance of the code in their areas. SIGs are chartered by the Steering Committee.

Working Groups (WGs) are **temporary groups** formed to address issues that cross SIG boundaries. Working groups do not own any code ownership or other long term artifacts. Working groups are chartered by the Steering Committee.

SIGs and WGs have open and transparent proceedings.

Anyone is welcome to participate and contribute provided they follow the Code of Conduct.

ONNX open governance update

Steering Committee

<https://github.com/onnx/steering-committee>

Prasanth Pulavarthi (MS)

Joe Spisak (FB)

Vin Sharma (AWS)

Harry Kim (Intel)

Dilip Sequeira (Nvidia)

Special Interest Groups (SIGs)

<https://github.com/onnx/sigs>

Architecture & Infra: Lu Fang, Ke Zhang

Operators: Michał Karzyński, Emad Barsoum

Converters: Chin Huang, Guenther Schmuelling

Model Zoo & Tutorials: Vinitra Swamy

Working Groups (WGs)

<https://github.com/onnx/working-groups>

Training: Svetlana Levitan

*Recently closed WGs: **Edge/Mobile, Data Pipelines***

ONNX open governance changes

Legal entity - LF AI is setting up a legal entity for ONNX. Charter is based on existing open governance plan.

Updated licensing - Project licensing is moving to Apache-2 (from MIT). Reclassification of prior contributions is under discussion.

CLA -> DCO - DCO is an attestation made by the contributor (versus a signed CLA) via the -s signoff (can be automated at contribution).

ONNX open governance election process

Now accepting applications until 4/20 (Mon): onnx.ai/sc-apply

Eligibility

Candidate: Self-nominated and not required to be a Contributor

Voter: Contributors to ONNX project

Methodology

- 1 vote per Member Company.
- Condorcet voting with Schultz method (Ranked preference)
- Contributor votes roll up to associated company
- Final result based on Member Company votes
- All votes published for transparency

Timeline

- April: Nomination and candidate campaigns
- May: Election and transition
- June: New Steering Committee

Questions?



ONNX

Wrap up!

Thank you ...

- Recording of today's workshop and other applicable content will be shared via ONNX-Announce mailing list when available.
- Please stay engaged and continue to contribute to ONNX and ONNX related projects.
- Remember to use the following ONNX resources:
 - Website: <https://onnx.ai/>
 - GitHub: <https://github.com/onnx>
 - Gitter: <https://gitter.com/onnx>
 - Calendar: <https://onnx.ai/calendar>
 - Mailing List: <https://lists.lfai.foundation/g/onnx-announce>