

ACM FCRC 2023

Essência das Palestras Plenárias

Daniel Rattton Figueiredo
PESC/COPPE/UFRJ



COPPE
UFRJ

ACM FCRC 2023



- ❑ ACM = Association for Computing Machinery
 - maior organização acadêmica e profissional da área de computação
 - promove congressos, revistas, currículos, biblioteca digital, políticas públicas, etc

ACM FCRC 2023

Federated Computing Research Conference

June 16-23, 2023 | Orlando World Marriott | Orlando, Florida

- ❑ FCRC: encontro que ocorre a cada 4 anos
 - juntar no mesmo local/data diferentes congressos da ACM (14 congressos em 2023)
 - palestrantes renomados de diferentes áreas
 - 2600 participantes (em 2023)

Profa. Margaret Martonosi



- ❑ Universidade de Princeton, Ciência da Computação
 - arquitetura de computadores, interface hardware-software
- ❑ Diretora da National Science Foundation (NSF) para Computer and Information Science and Engineering (CISE)
 - orçamento anual de mais de US \$1B para investimento em pesquisa

Taking on the World's Challenges

□ The Role of Computing Research & Innovation

Societal Grand Challenges



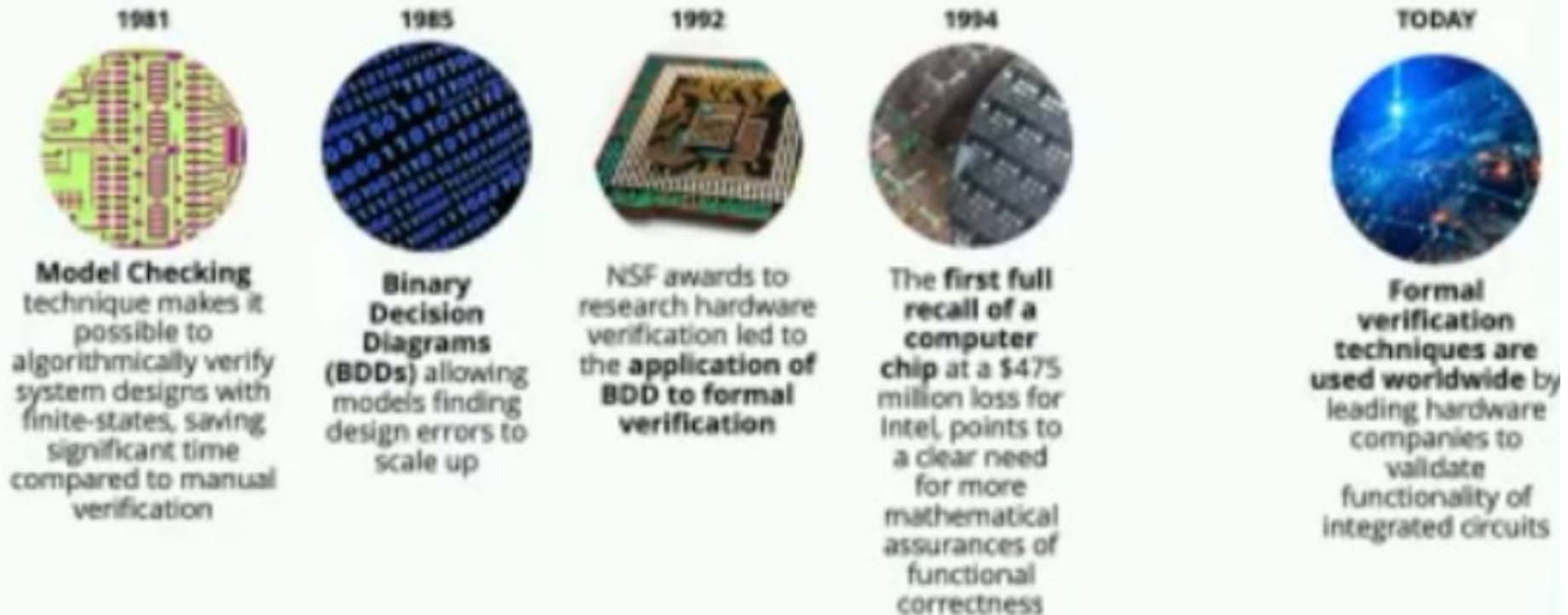
The infographic displays three societal grand challenges. Each challenge is represented by a circular icon above a colored rectangular box containing the challenge name and a question. The challenges are: 1. Climate & Sustainability (dark blue box), 2. Trust in Information (gold box), and 3. People (teal box). The NSF logo is in the bottom left corner, and a small number '3' is in the bottom right corner.

| Challenge | Icon Description | Question |
|--------------------------|---|---|
| CLIMATE & SUSTAINABILITY | Icon showing a person, a tree, and a flame. | How might we expand the frontiers of knowledge and technology for a sustainable future? |
| TRUST IN INFORMATION | Icon showing two hands shaking. | How best to provide trustworthy and meaningful access to information. What is truth? |
| PEOPLE | Icon showing a group of people. | How might we be more inclusive in engaging the world's full talents in our work? |

□ Foco em projetos passados e futuros da NSF

Histórias de Impacto

Impact story: NSF-funded Research in Formal Verification



- ❑ Investimento em pesquisa de base dando origem a tecnologia de ponta 40 anos depois!
 - importância do investimento em pesquisa

Profa. Shafi Goldwasser



- ❑ Universidade da Califórnia Berkeley, Eng. Elétrica e Ciência da Computação
 - teoria da computação, criptografia
 - ACM Turing Award (2012), Gödel Prize (1993)
- ❑ Diretora do Simons Institute for the Theory of Computing

Constructing and Deconstructing Trust

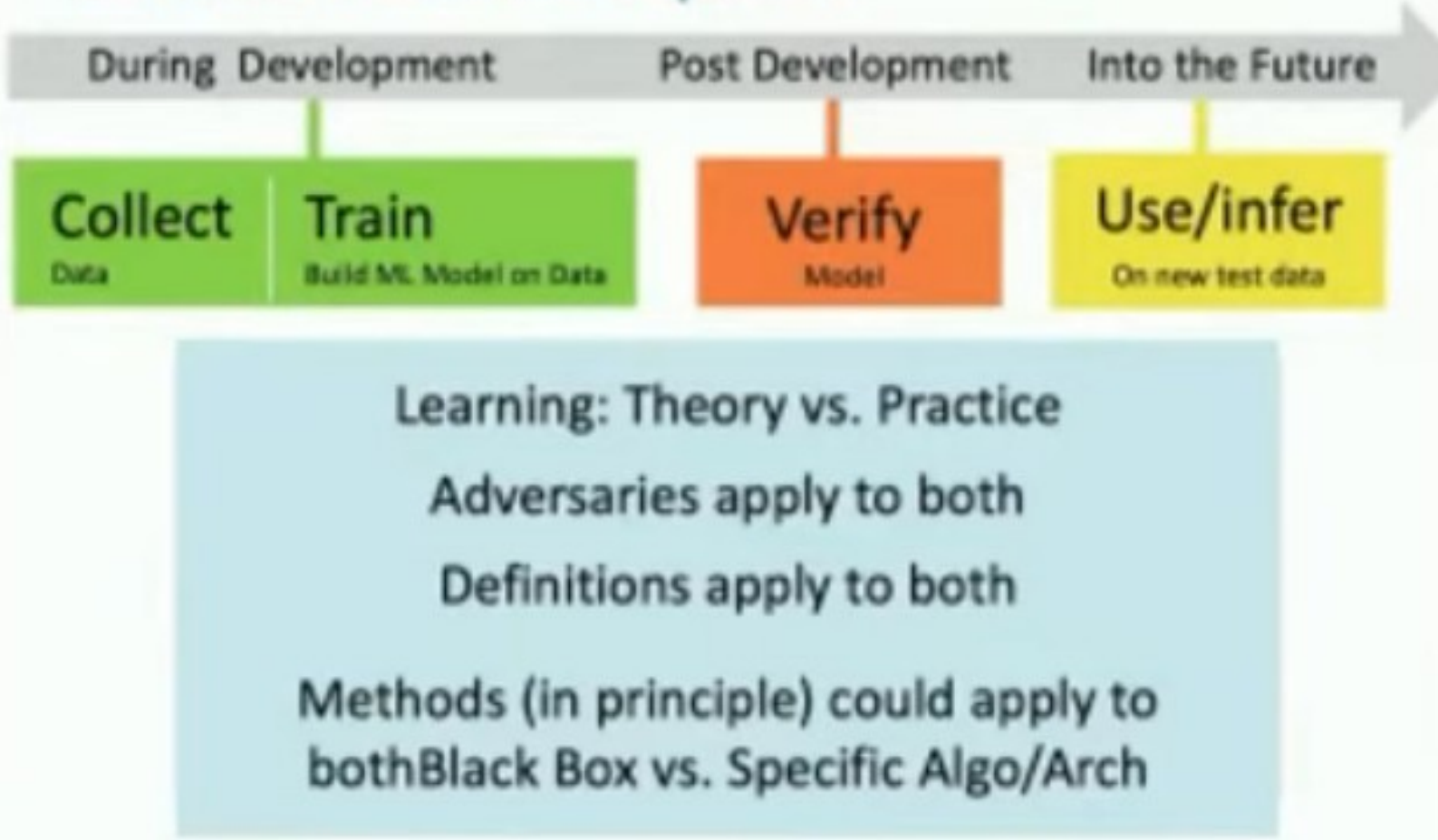
- ❑ Employing Cryptographic Recipe in the ML Domain
- ❑ Crise de (falta de) confiança em sistemas de IA
- ❑ Confiança em sistemas criptográficos
- ❑ Trazer para IA paradigma de construção de confiança

ML/AI was **NOT** originally designed for Adversarial Contexts

- ~~Not Integral Part of the Definition of the Problem~~
- And yet AI systems are VERY attractive targets
- **Adversarial modeling:** key to safe usage and composability
 - Do not make assumption on the Adversary Strategy – prepare for **worst case**
 - Do assume computational limits on adversary time.

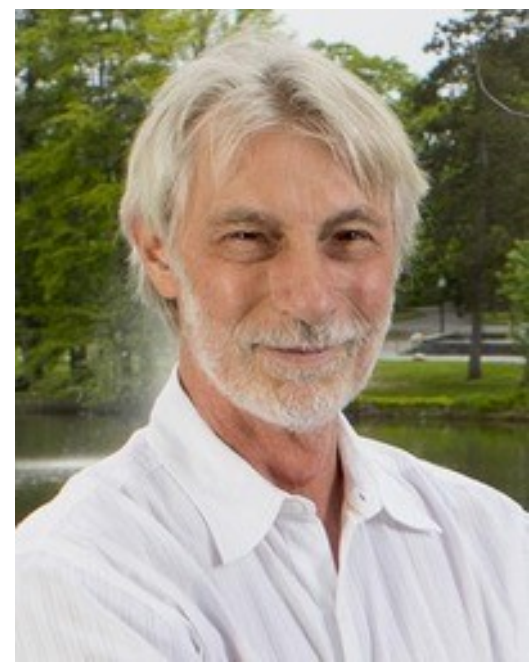
Adversários e Estágios

Adversaries in ML Pipeline



- ❑ Considerar modelos para o adversário em diferentes etapas do processo de aprendizado
 - muitos trabalhos recentes nesta área

Prof. Don Towsley



- ❑ Universidade de Massachusetts Amherst, Ciência da Computação
 - redes, performance
- ❑ Achievement awards: 2007 ACM Sigmetrics, 2008 ACM Sigcomm, 2011 IEEE Infocom
- ❑ Membro correspondente da Acad. Bras. de Ciências
 - orientou o doutorado de 4 brasileiros, orienta 1 agora (ex-ECI)

The Quantum Internet

Recent Advances and Challenges

The Quantum Internet

Vision: Quantum network enabling full quantum connectivity between multiple user groups.

Secure Communications

Quantum Multi-User Applications

Sensing, Timing, GPS

Networked Quantum Computing

QUANTUM SWITCH (QS)

QUANTUM REPEATER (QR)

QUANTUM ROUTER

QUANTUM COMPUTER (QC)

USER

acm FCRC

The slide features a central network diagram with nodes and connections. A speaker is visible on the right side of the stage. The background is a dark blue with a starry pattern.

Quantum Internet?

Why Quantum Internet?

Cryptography, security – quantum key distribution (QKD)

Distributed quantum computing – breaking web security, solving hard problems

High resolution sensing – exploring the universe

Teleportation of quantum data



Why don't we have a quantum Internet yet?

It's hard!

Why?

Loss & noise

- ❑ Técnicas para lidar com perda e ruídos!

Prof. Torsten Hoefler



- ❑ Universidade ETH Zurich (Suíça), Ciência da Computação
 - sistemas de alto desempenho
- ❑ Prêmios de melhor artigo: ACM/IEEE Supercomputing Conference SC10, SC13, SC14, SC19, SC22, EuroMPI'13, HPDC'15, HPDC'16, IPDPS'15
- ❑ Mais jovem dos palestrantes, único atuando fora dos EUA

Scalable and Efficient AI

From Supercomputers to Smartphones

Supercomputers fuel Modern AI

Facebook's parent, Meta, creates powerful AI supercomputer

Google artificial intelligence supercomputer creates its own 'AI child' that can outperform its human-made rivals

Microsoft invests \$1 billion in OpenAI to pursue holy grail of artificial intelligence

Tesla unveils Dojo supercomputer: world's new most powerful AI training machine

AI SUPERCOMPUTER
10k GPUs

reddit
A robot may ___ injure a human being or, through inaction, allow a human being to come to harm.

layer-wise weight update

layer-wise weight update

| | | | |
|-----------|------|-----------|------|
| not | 0.74 | not | 0.00 |
| sometimes | 0.26 | sometimes | 0.00 |
| always | 0.01 | always | 0.00 |
| never | 0.04 | never | 0.00 |
| and | 0.01 | and | 0.00 |
| but | 0.02 | but | 0.00 |
| never | 0.02 | never | 0.00 |

- GPT-3: 500 billion tokens
- ImageNet (22k): A few TB
- Soon: **the whole internet!**
- GPT-3: 96 [complex] layers
- 175 bn parameters (**700 GiB** in fp32)
- 2048-token "sentences"

See: Mar. 21. Democratizing our AI and distributed deep learning: An in-depth research analysis, ACM Computing Surveys, 55(4), 2021

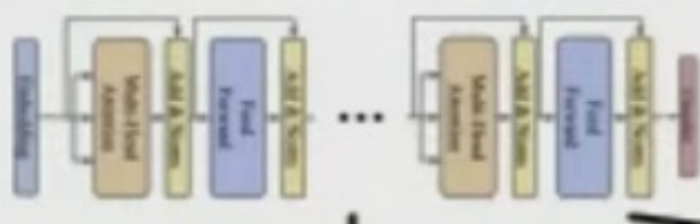
Empresas adquirindo supercomputadores para ajustar (treinar) modelos gigantescos!

Scalable and Efficient AI

- Três dimensões para reduzir custo computacional
 - entrada/saída, calcular, comunicar

SPCL ETH zürich

Three Systems Dimensions in Large-scale Super-learning ...



High-Performance I/O

- Quickly growing data volumes
 - Scientific computing!
- Use the specifics of machine learning workloads
 - E.g., intelligent prefetching

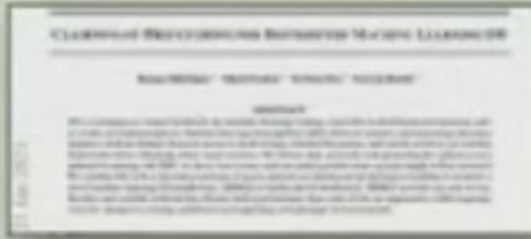

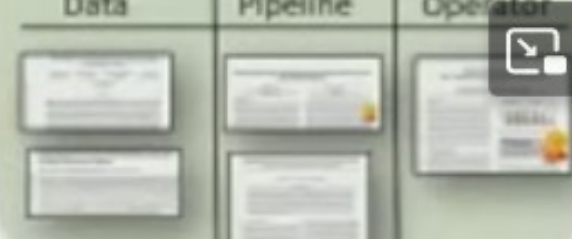
High-Performance Compute

- Deep learning is HPC
 - Data movement!
- Quantization, Sparsification
 - Drives modern accelerators!

High-Performance Communication

- Use larger clusters (10k+ GPUs)
- Model parallelism
 - Complex pipeline schemes
- Optimized networks

Distribution and Parallelism

| Data | Pipeline | Operator |
|--|--|---|
|  |  |  |

- Fundamental para permitir outros *players*
 - permite ainda mais dados e maiores modelos

Fim

- ❑ Biografia dos palestrantes e resumo das palestras
 - <https://fcrc.acm.org/program/plenary-speakers>
- ❑ Palestras gravadas (com perguntas ao final)
 - <https://www.youtube.com/playlist?list=PLn0nrSd4xjjZ5DcBqu8xxoFQ3QjqAaQXI>



- ❑ Perguntas, dúvidas, comentários?