HYPERION RESEARCH

The Need for Deep Learning Transparency

with Speaker Notes

Steve Conway, SVP-Research

March 2018

Stephen Hawking on AI

"Success in creating effective AI could be the biggest event in the history of our civilization…[But] unless we learn how to prepare for and avoid the potential risks, AI could be the worst event in the history of our civilization.“

November 2017

HYPERION RESEARCH

Hyperion Definitions

AI: Machine Learning, Deep Learning

"We still don't have a universally accepted definition of what intelligence is, so it would be hard to [define] artificial intelligence." Larry Greenemeier, Associate Editor for Technlogy Scientific American (March 2018)

- **Artificial Intelligence (AI):** a broad, general term for the ability of computers to do things human thinking does (but NOT to think in the same way humans think). AI includes machine learning, deep learning and other methodologies.
- **Machine learning (ML)**: a process where examples are used to train computers to recognize specified patterns, such as human blue eyes or numerical patterns indicating fraud. The computers are unable to learn beyond their training and human oversight is needed in the recognition process. The computer follows the base rules given to it.
- **Deep Learning (DL):** an advanced form of machine learning that uses digital neural networks to enable a computer to go beyond its training and learn on its own, without additional explicit programming or human oversight. The computer develops its own rules.

Forecast:
HPDA Market and ML/DL/AI Methods

TABLE 1

Worldwide HPC AI Server Revenues vs. All HPDA Server Revenues ($ Millions)

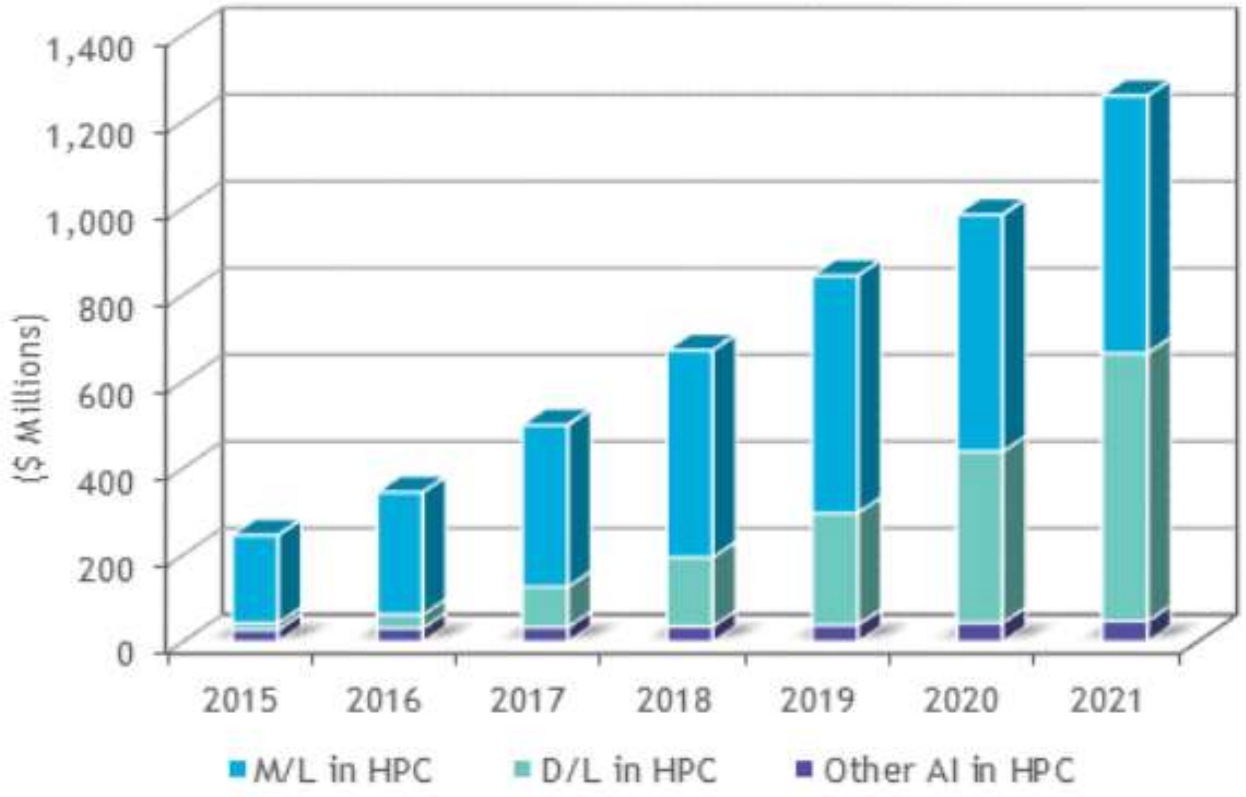
	2015	2016	2017	2018	2019	2020	2021	CAGR 16-21
Total WW HPDA Server Revenues	$1,455	$1,845	$2,333	$2,830	$3,224	$3,488	$4,040	17.0%
Total HPC-Based AI (DL, ML, and Other)	$246	$346	$501	$673	$845	$986	$1,260	29.5%

Source: Hyperion Research 2017

WW M/L, D/L, & AI Forecasts

FIGURE 2

Worldwide M/L, D/L & AI HPC-Based Revenues



Source: Hyperion Research 2017

AI/Deep Learning Major Challenges

"The amount of data available today is miniscule compared to what we need for deep learning."

Marti Head, GlaxoSmithKline

MARKET STATUS

- HPC has moved to the forefront of DL/AI research
- Ecosystem (including GPGPUs) formed around social media/Web giants
- DL needs massive data: not available yet in many markets
- Lack of standard benchmarks lengthens sales process
- Need for transparency ➡ HPC simulation!
- Lots of time/*money* being spent to get there

The AI/Deep Learning (DL) Challenge

- We can't teach machines to think like humans, because we don't fully understand how humans think.
- DL machines can learn on their own -- beyond the instructions humans give them.
- DL machines are capable of learning from each other, i.e., they are capable of culture*

  * *Aggregation of knowledge among individuals over time*
- Today, DL learning is largely opaque to humans -- the basis for DL inferences/ decisions is unclear.
- DL "black boxes" need to be made transparent.

The DL Transparency Challenge Is Pervasive

"Deep learning methods are fairly opaque to human inspection… intelligence analysts are unlikely to trust a system unless they understand how its results are achieved." **IARPA**

"Deep neural networks are notoriously opaque. If consumers are to entrust their safety to AI-driven vehicles or their health to AI-assisted medical care, they will want to know how these systems make critical decisions." **University of Oxford Future of Humanity Institute**

"The design of the [supercomputer for oncology] system is such that it cannot explain why it weighs one treatment over another." **Stat report on AI-driven system at Memorial Sloan-Kettering Hospital**

DL medicine "is inherently opaque." **Nicholas Price, Univ. of Michigan health law scholar**

Game

Google's DeepMind (AlphaGo) defeats the best humans.

"We still can't explain it…you could…review…every parameter in AlphaGo's artificial brain, but even a programmer would not glean much from these numbers because what drives a neural net to make a decision is encoded in the billions of diffuse connections between nodes."

Alan Whitfield, robot ethicist, Univ. of the West of England

Life

Self-Driving Uber Vehicle Strikes and Kills Pedestrian

Washington Post, March 19, 2018

Automakers, insurance companies & families need to know why.

Game

IBM Watson defeats best humans at Jeopardy! game show.

"I felt like 'Quiz Show Contestant' was now the first job that had become obsolete under this new regime of thinking computers."

Ken Jennings, game show champion who lost to Watson

Life

[The oncology supercomputer at Memorial Sloan Kettering Hospital] sometimes suggests a chemotherapy drug for patients whose cancer has spread to the lymph nodes, even when it has been given information about a patient whose cancer has ***not*** spread to the lymph nodes.

Slate.com, Septermber 6, 2017

NHTSA Report: At Minimum, There Must Be After-the-Fact Transparency

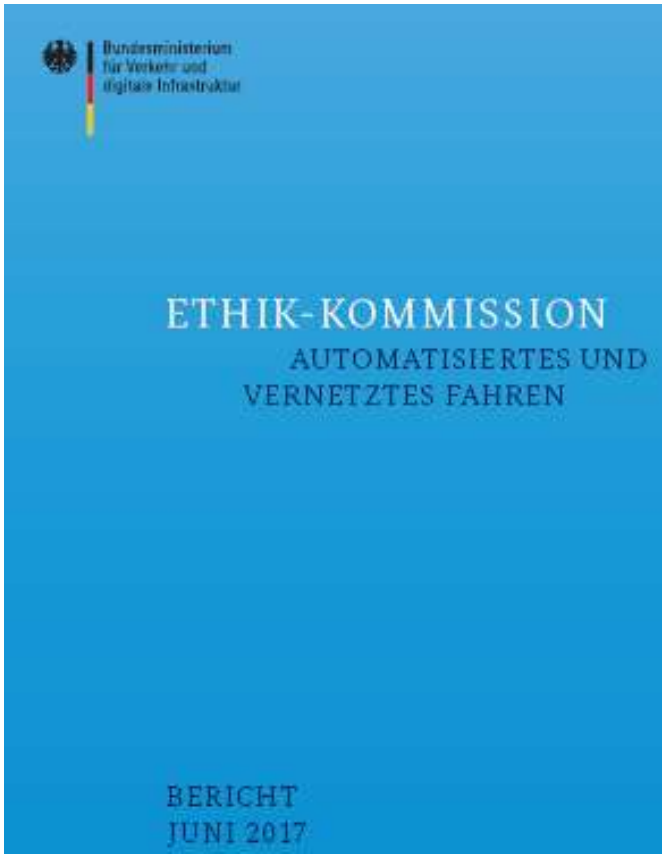
National Highway Transportation Safety Alliance

"Vehicles should record…all available information relevant to the crash, so [it] can be reconstructed."

"No standard data elements exist…to use in determining why an ADS-enabled vehicle crashed."

June 2017: Germany Passes First Law Governing Autonomous Vehicles

- 14-person Ethics Commission led by Transportation Minister
- All cars operating in Germany must let humans take control.
- Accidents:
  - If car's in control, automaker is liable
  - If person's in control, person is liable
- Cars can't be programmed demographically:
  - E.g., allow an elderly person to die before a baby
- March 2018: China invites member of German Ethics Commission to advise on China's ADS law.



Report: "Autonomous and Networked Driving"

Approaches to the DL Transparency Problem

- **Manual testing**
  - Humans feed test images into the network until they spark a wrong decision
- **Adversarial testing**
  - Computer automatically tweaks a specific image until it causes a wrong decision
- **Deep Xplore**
  - Assumes neural networks usually make the right decision.
  - Polls 3 or more neural networks and retrains the dissenting network to conform to the majority.

"The goal of DARPA's Explainable Artificial Intelligence (XIA) program is make AI-controlled systems more accountable to their human users."

NERSC
NATIONAL ENERGY RESEARCH SCIENTIFIC COMPUTING CENTER

COLUMBIA UNIVERSITY

LEHIGH UNIVERSITY

Another DL Issue: "Catastrophic Forgetting"

- When a neural network encounters something it wasn't trained to recognize, it tends to make the same mistake over and over again.
- Attempts to retrain the network on the fly lead to "catastrophic forgetting": learning the new thing disrupts the device's prior knowledge.
  - With humans, moving a tennis net one foot wouldn't require re-learning the game of tennis.

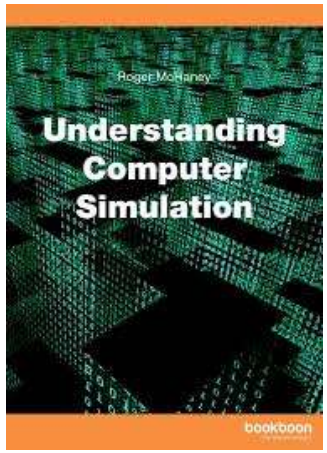
Working on it…

neurala

ADS Traffic: Multiple Control Levels

HPC Simulation Is Not Going Away Any Time Soon

- Neural networks are not good today at simulating physics under varying conditions.
- Even if they improved, the transparency issue would make if difficult to validate scientific use cases.



"There are many things we have been getting right in the simulation community based on 30-40 years of applied math, and so a black box, as good as it may be, will not be suitable until these things are worked out."

NERSC's Prabhat. NextPlatform, 1/26/18

HYPERION RESEARCH

QUESTIONS?

ejoseph@hyperionres.com

sconway@hyperionres.com

bsorensen@hyperionres.com

anorton@hyperionres.com

jeansorensen@hyperionres.com

mthorp@hyperionres.com

kgantrish@hyperionres.com