AIDC

INTEL AI DEVCON 2018

SAN FRANCISCO | MAY 23-24

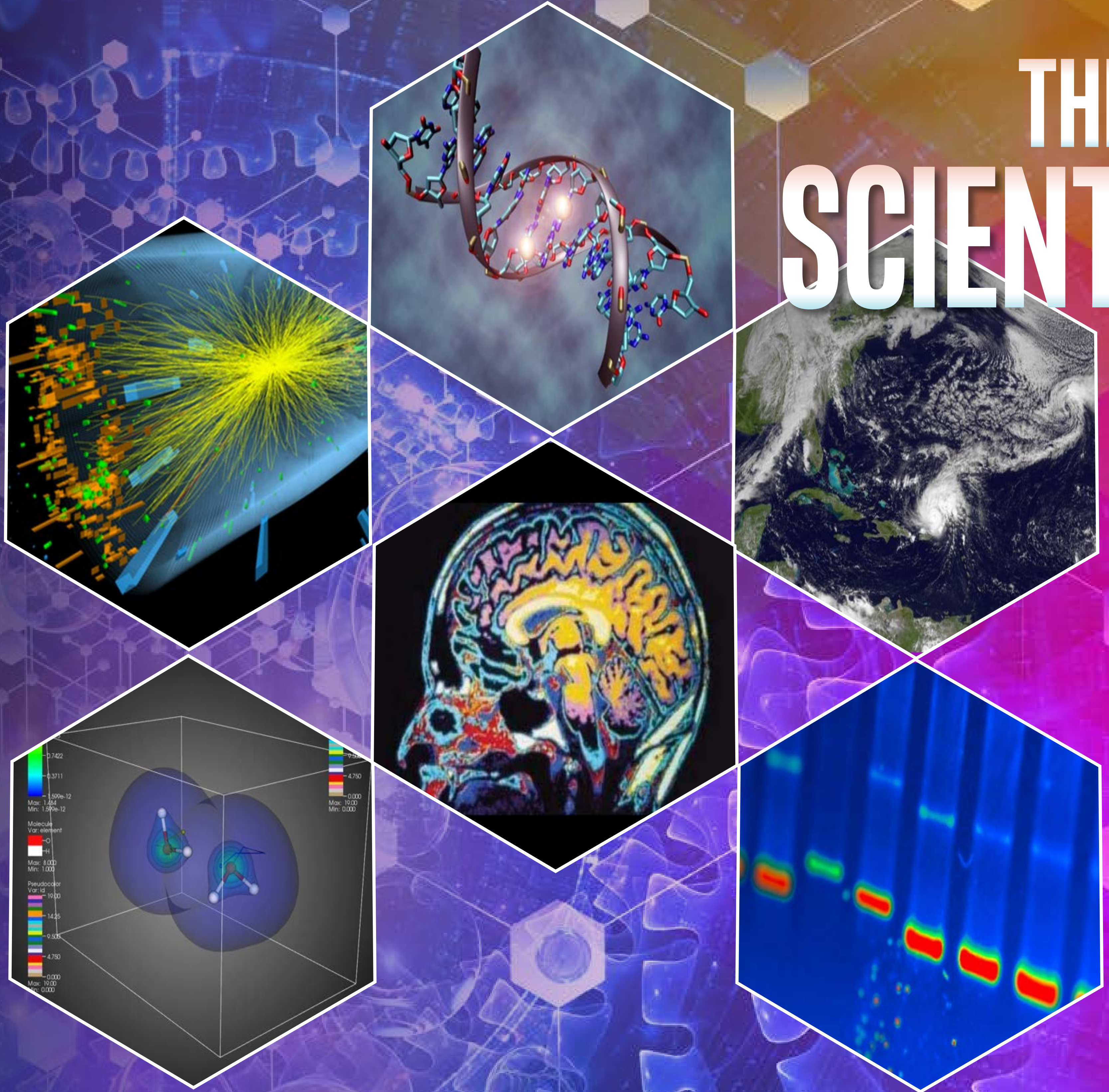
AIDC
INTEL AI DEVCON 2018

DEEP LEARNING AND THE TRANSFORMATION OF SCIENTIFIC EXPLORATION

GADI SINGER

VICE PRESIDENT,
ARCHITECTURE GENERAL MANAGER,
ARTIFICIAL INTELLIGENCE PRODUCTS GROUP
INTEL CORPORATION

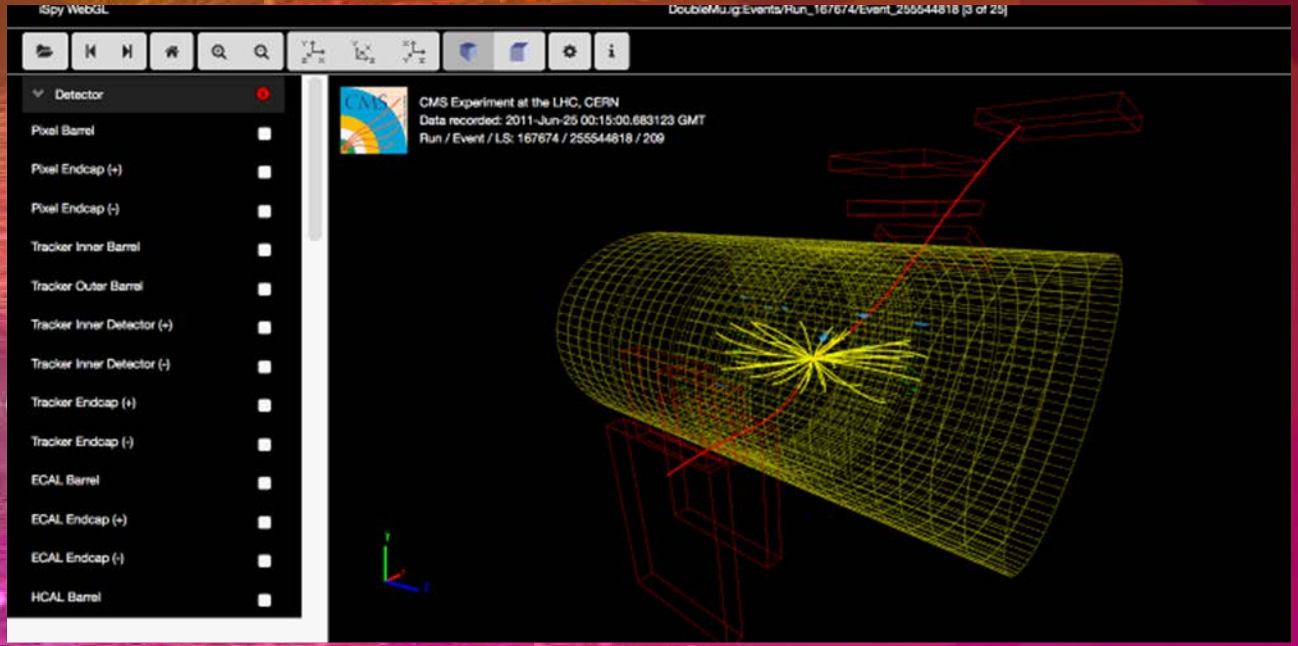
THE RISE OF AI-ENRICHED SCIENTIFIC EXPLORATION

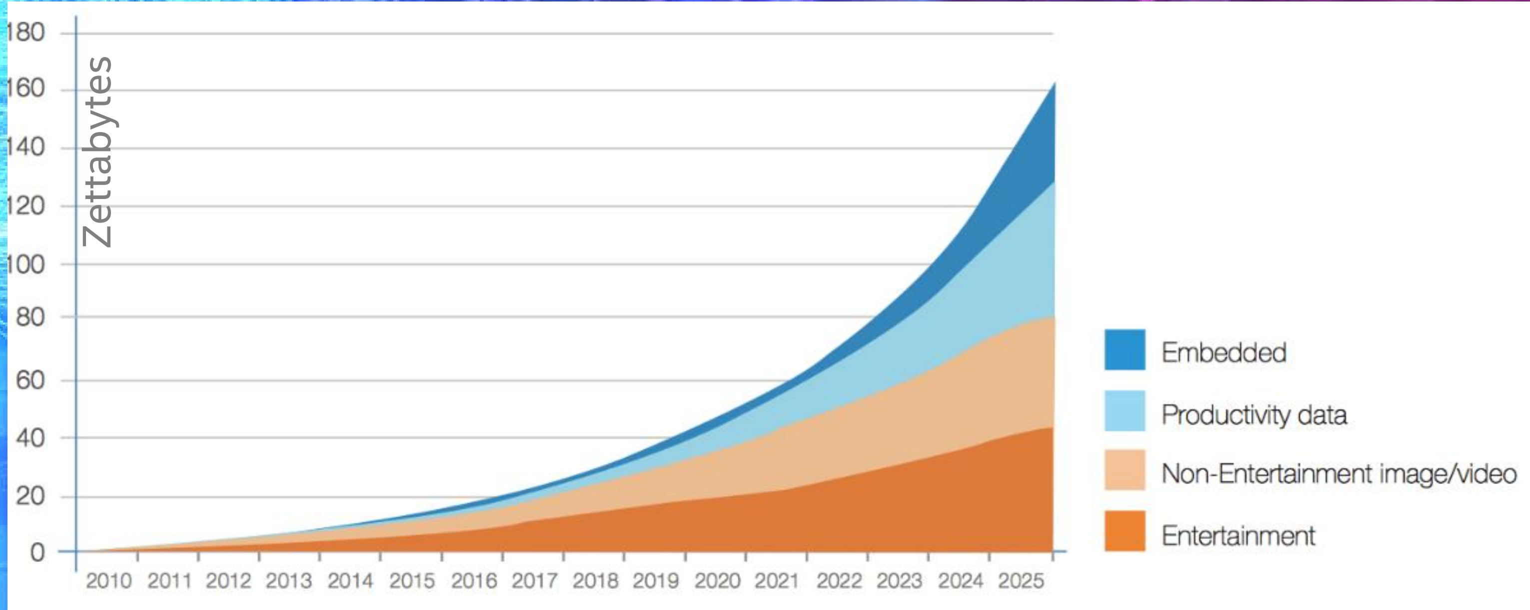


SCIENTIFIC PARADIGMS

- EMPIRICAL
- THEORETICAL
- COMPUTATIONAL
- DATA EXPLORATION (PROGRAMMED)
- PHENOMENON EXPLORATION (ML)

THE CAMBRIAN EXPLOSION OF DATA

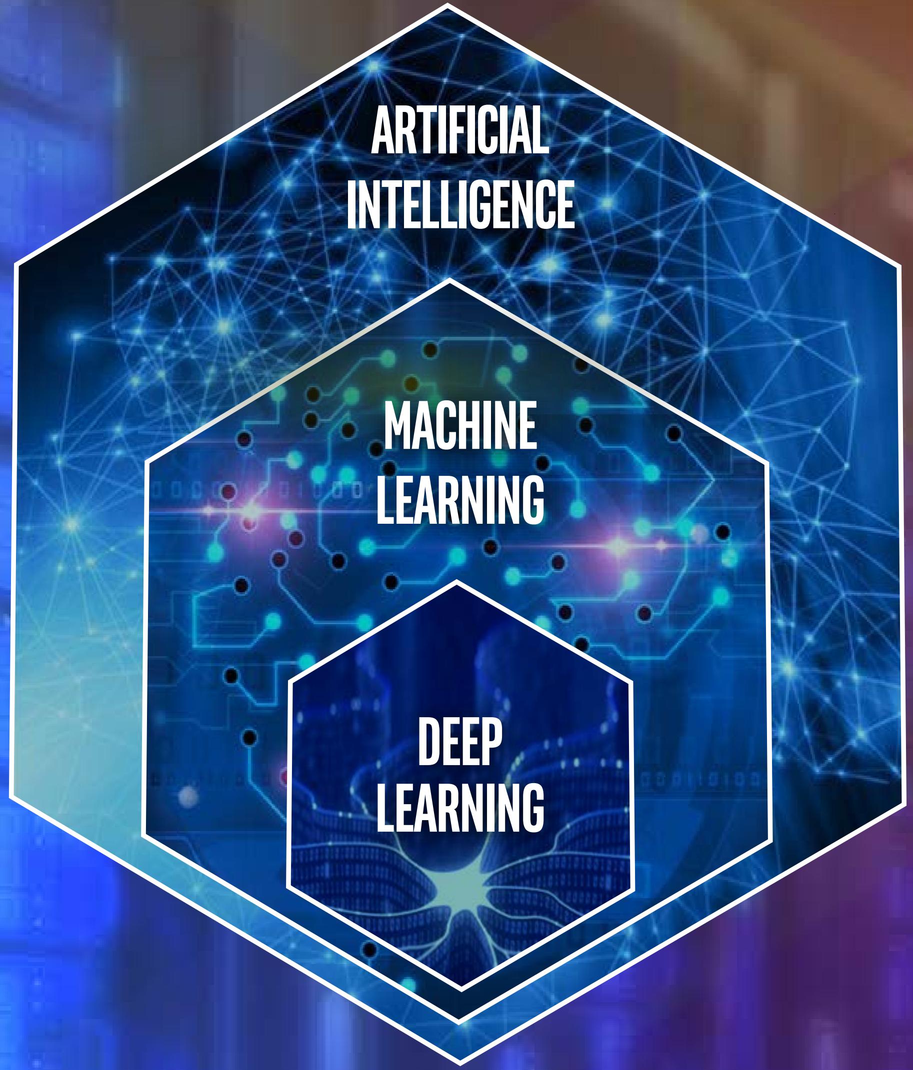




Source: IDC's Data Age study, sponsored by Seagate, April 2017

	DAILY BY 2020
AVERAGE INTERNET USER	1.5 GB
AUTONOMOUS VEHICLE	4 TB
CONNECTED AIRPLANE	5 TB
SMART FACTORY	1 PB
CLOUD VIDEO PROVIDER	750 PB

ZETTA DATA × EXA COMPUTING × MACHINE LEARNING



DEEP LEARNING IN HPC AND SCIENCE

SPEECH

VISION

ADAS

FRONT FACING CAMERAS

REAR FACING CAMERA

LATERAL CAMERA

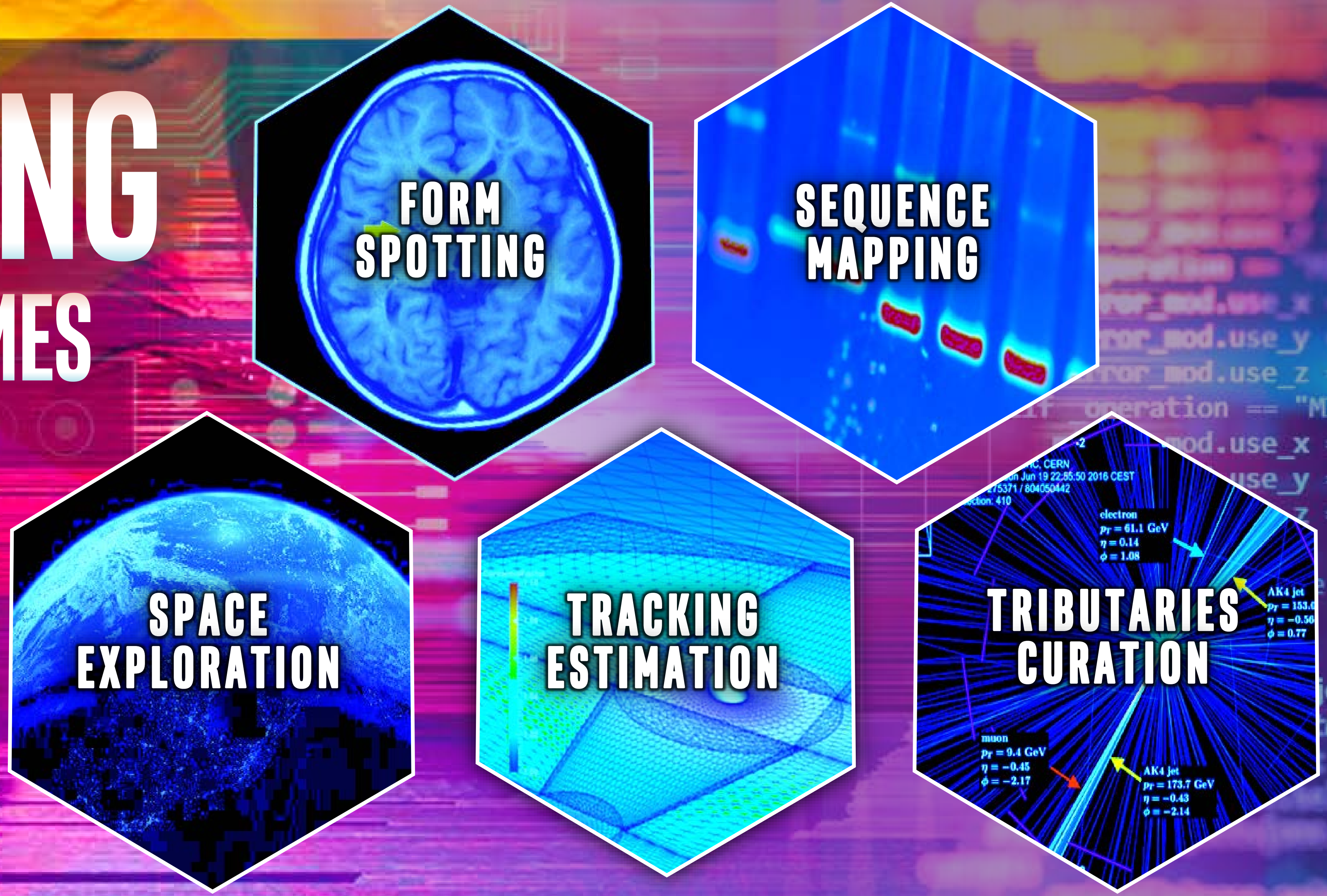
ALPHA GO

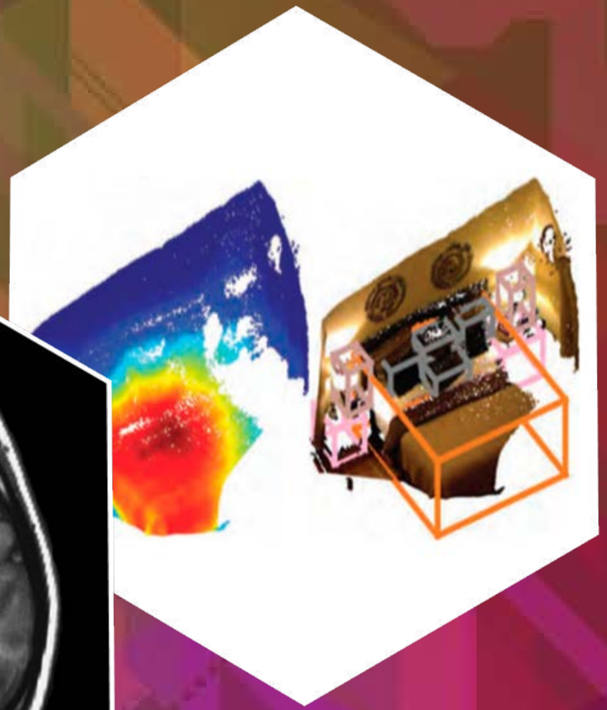
TYPES OF ANALYTICS/ML (PARTIAL LIST)

- CLASSIFICATION
- REGRESSION
- CLUSTERING
- FEATURE LEARNING
- ANOMALY DETECTION

MACHINE LEARNING

CLASSES OF APPLICATION SCHEMES





FORM SPOTTING

FINDING A BLADE OF GRASS IN A HAYSTACK

- Finding a ‘kind of pattern’ in multi-dimensional elaborate data
- Lots of examples available
- Weak signal in a sea of noise
- No mathematical/statistical model

Applicable DL techniques:

Pattern classification

Feature learning

Anomaly detection

Supervised learning:

Data tagging

FORM SPOTTING
NEUROSCIENCE

PRINCETON
Neuroscience
INSTITUTE

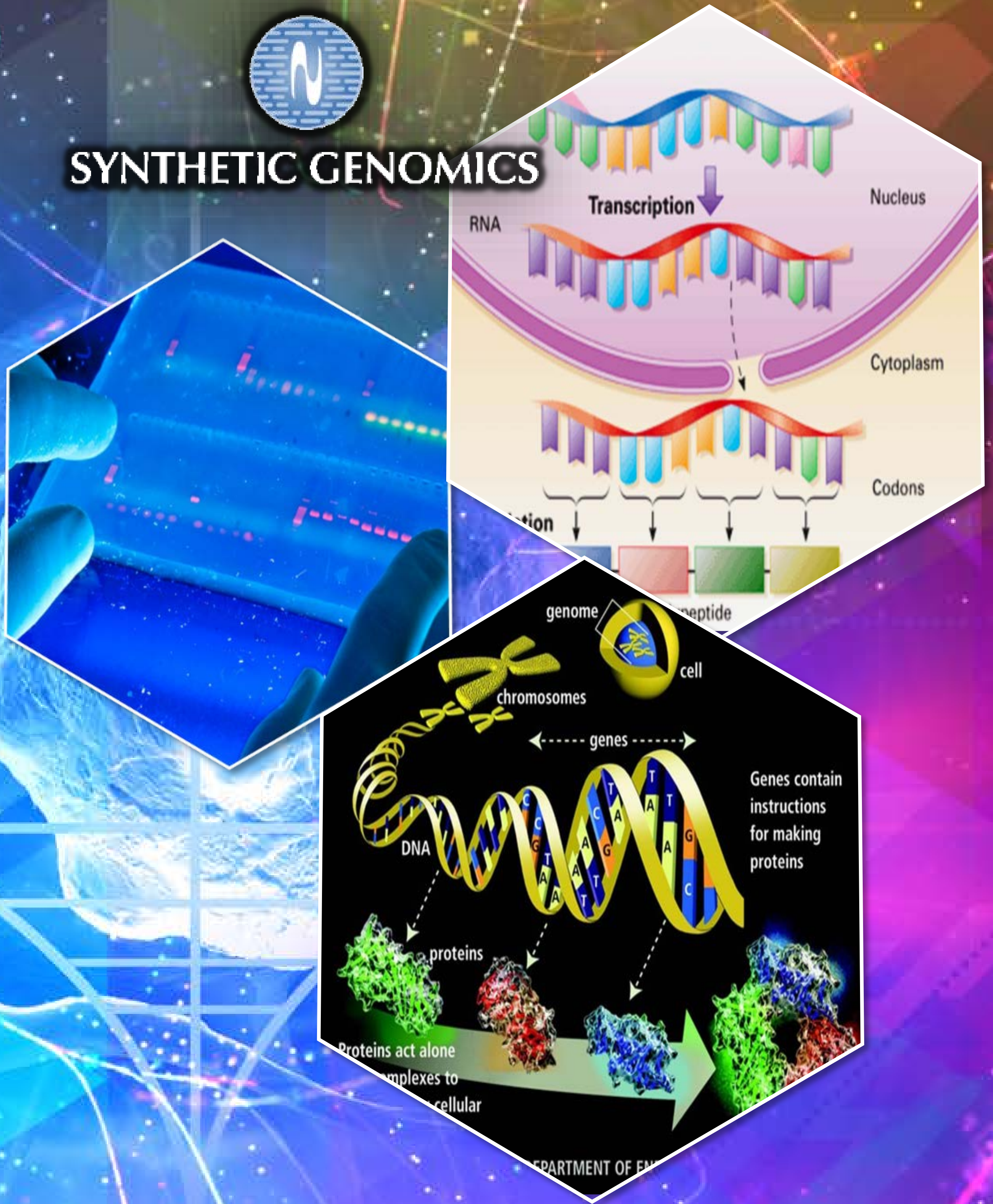
http://brainiak.org

Princeton Neuroscience Institute mapped the human mind in real time for improved diagnosis and treatment of brain disorders and mental illness.

Typical single scan (~1 million voxels) evaluated in seconds vs hours.

BrainIAK - Developing the next generation in fMRI brain imaging.

SYNTHETIC GENOMICS



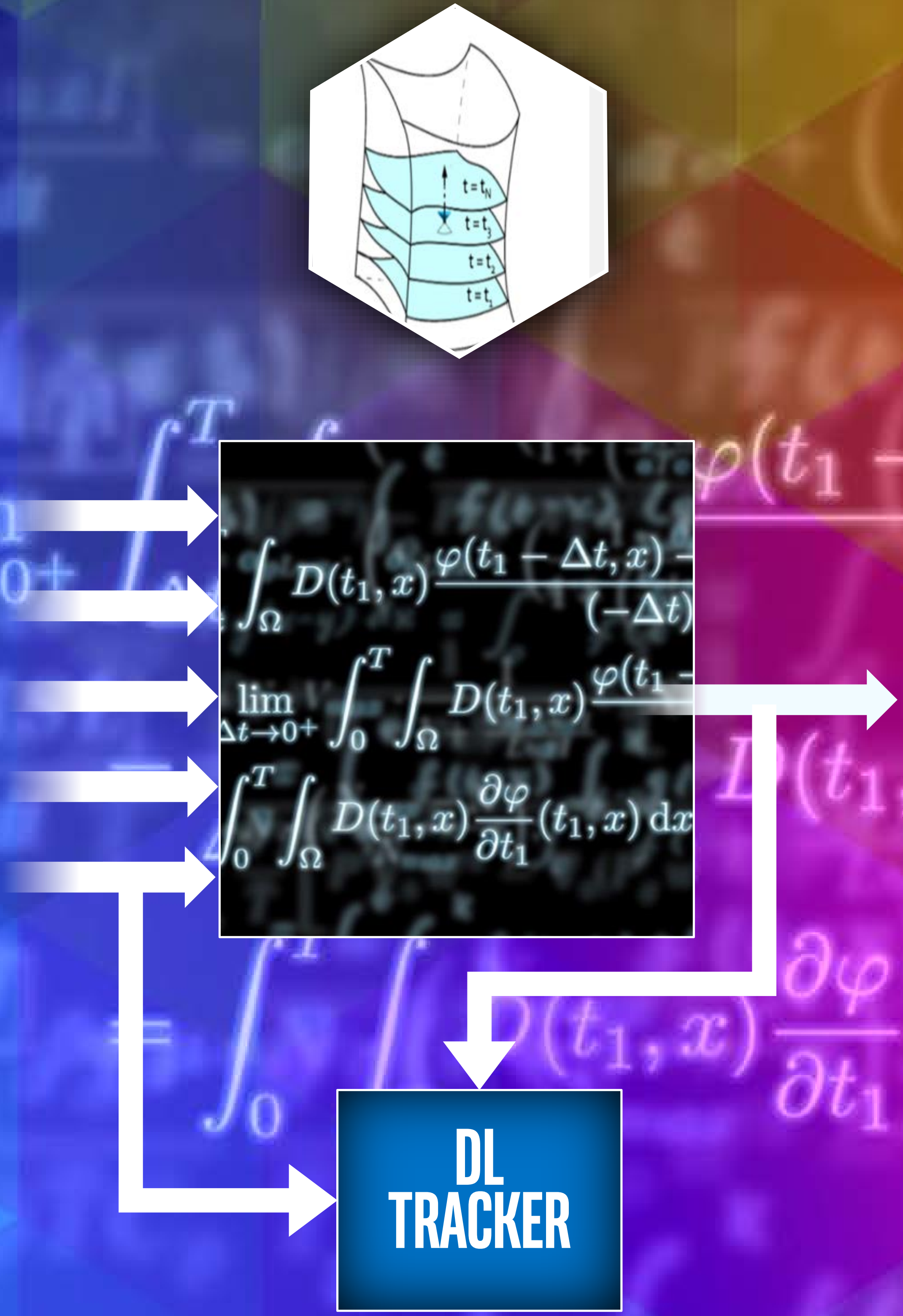
FORM SPOTTING

GENOMICS

- 10^4x speed up boost: annotate 1 million genes in <1hr vs. weeks with traditional tools
- Assign function to millions of uncharacterized proteins
- Semantic Search: discover proteins with related function even without sequence similarity
- Early stages of protein design: predict in seconds impact of every possible AA change

Joint effort of SGI and Intel

TRACKING ESTIMATION



- Well defined functions/model; compute intensive
- Major reduction (e.g., 10^4 x) required to enable real-time or rapid iterations

Applicable DL techniques:

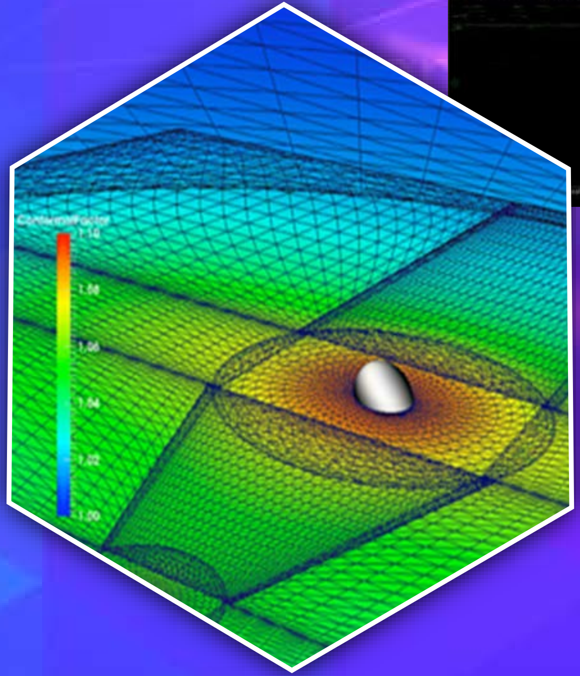
Regression

Supervised learning:

Full model training the DL Shadowing Estimator

TRACKING ESTIMATION

DETECTION OF GRAVITATIONAL WAVES (LIGO)



- Laser Interferometer Gravitational-Wave Observatory (LIGO) labs
- Detection of gravitational waves from binary black hole mergers
- Process array of sensors for directing a high-focus radio telescope
- Real-time multimessenger detection (DNN) >10^4 speedup: multiple days to 'real-time'
  (George, D. , Huerta, E. A.: *Deep Neural Networks to Enable Real-time Multimessenger Astrophysics*)

ILLINOIS

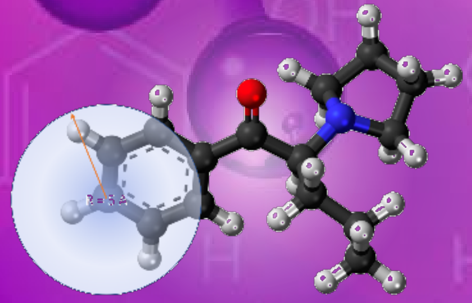
LIGO

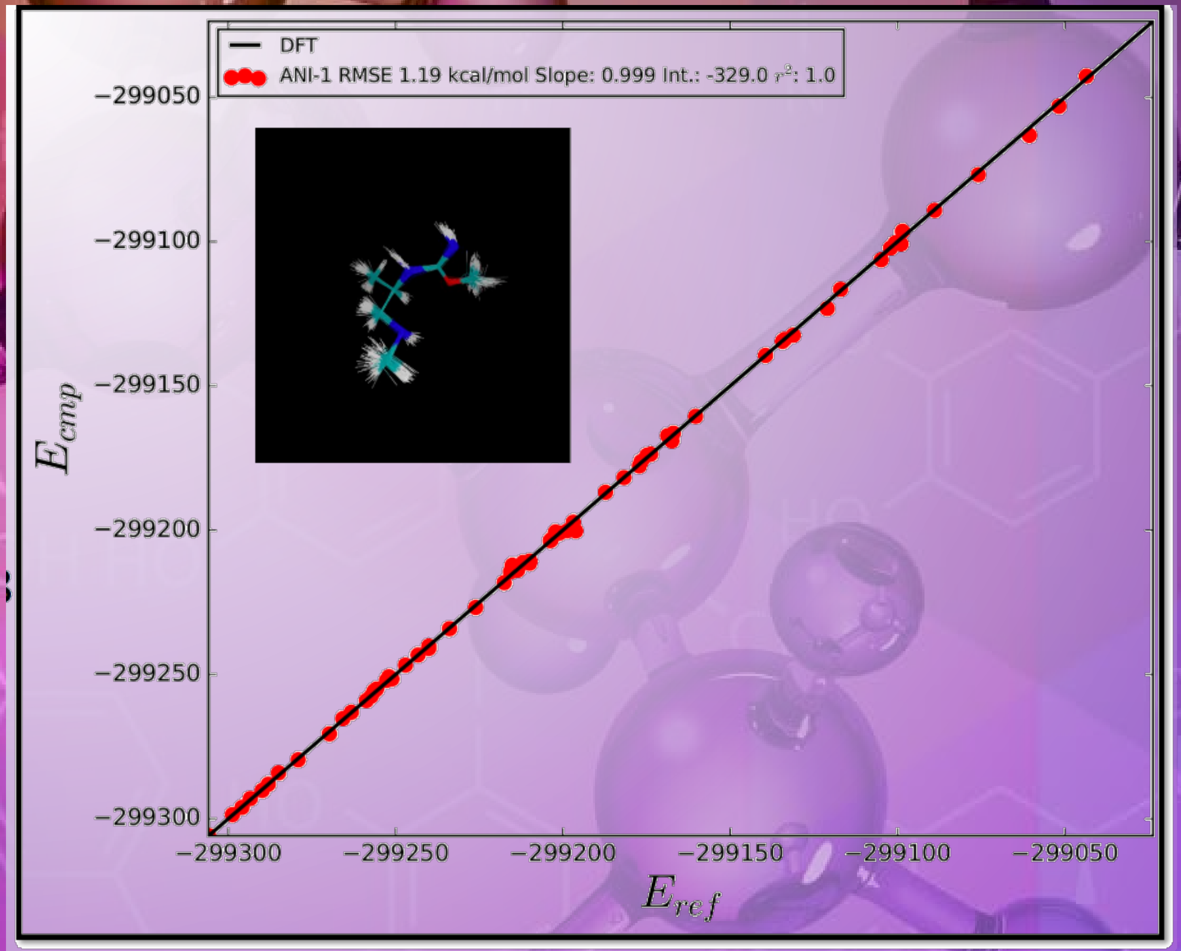
https://www.ligo.caltech.edu

TRACKING ESTIMATION

APPROXIMATING THE BEHAVIOR OF MOLECULES

- Predicting behavior of organic molecules
- Compute intensive Kohn-Sham Density-Functional Theory (DFT) equations
- Database of 20 million conformations
- Chemically accurate DL
- 10^5 speedup; ~$6x10^{-4}$ power reduction





Source: Mastering Computational Chemistry with Deep Learning, Isayev, O, University of North Carolina Chapel Hill

UNC ESHELMAN SCHOOL OF PHARMACY

SEQUENCE MAPPING

- Creating output sequence based on context based, multi-dimensional, continuous input sequence

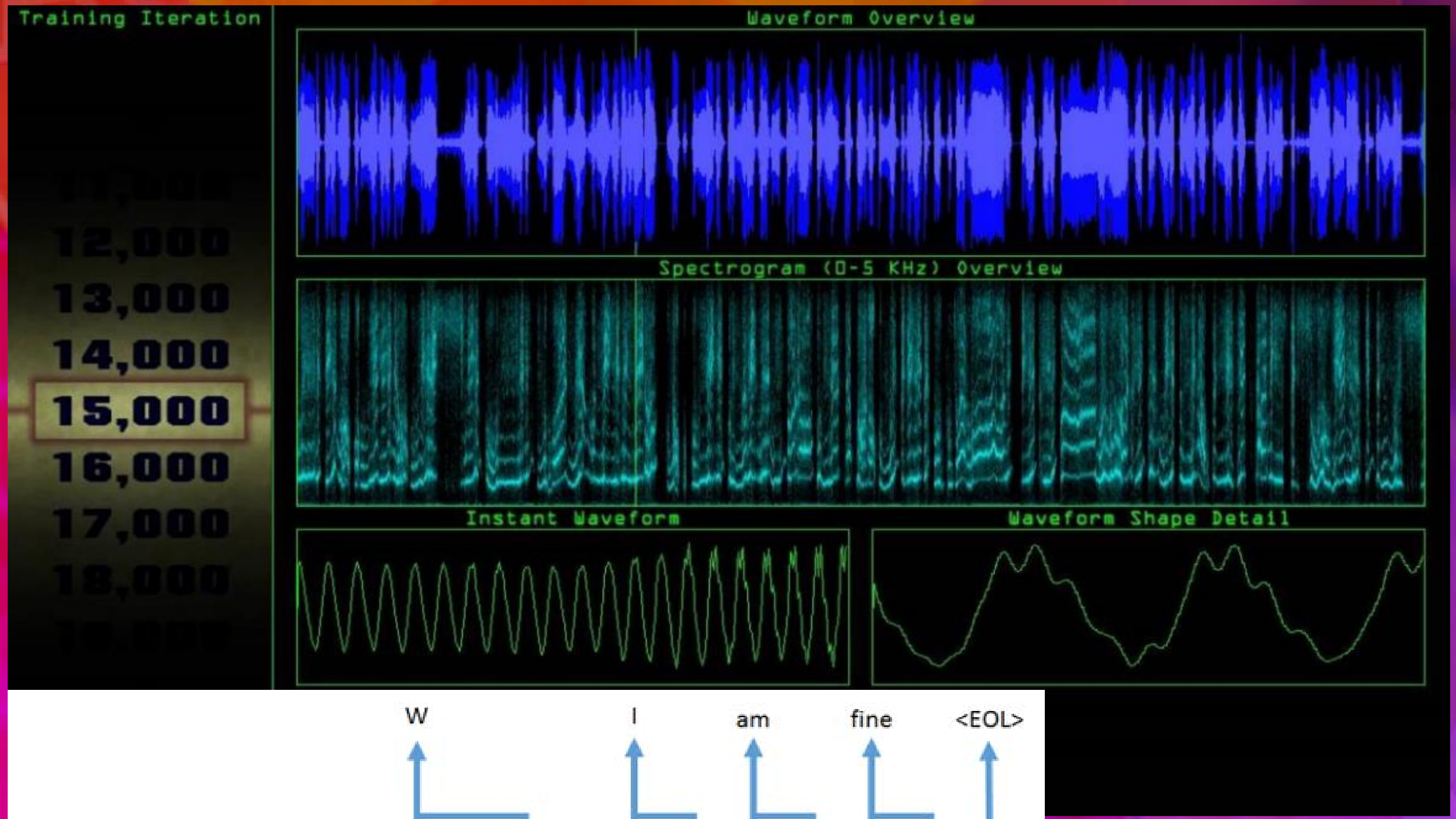
Applicable DL techniques:

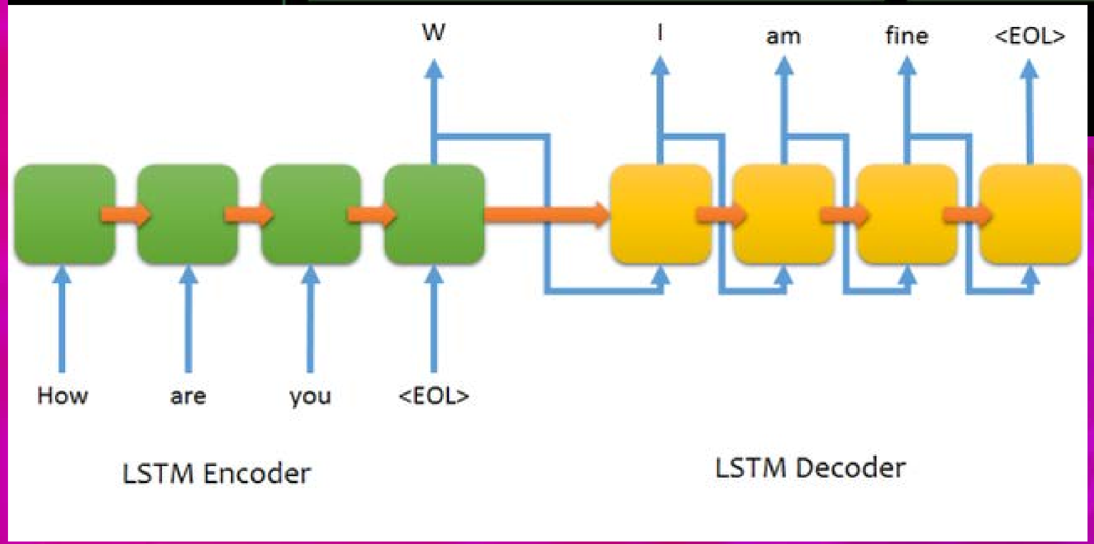
Neural Machine Translation (NMT)

Sequence-to-sequence transformation

Supervised learning:

Sequence examples tagging





AIDC
INTEL AI DEVCON 2018

SEQUENCE MAPPING

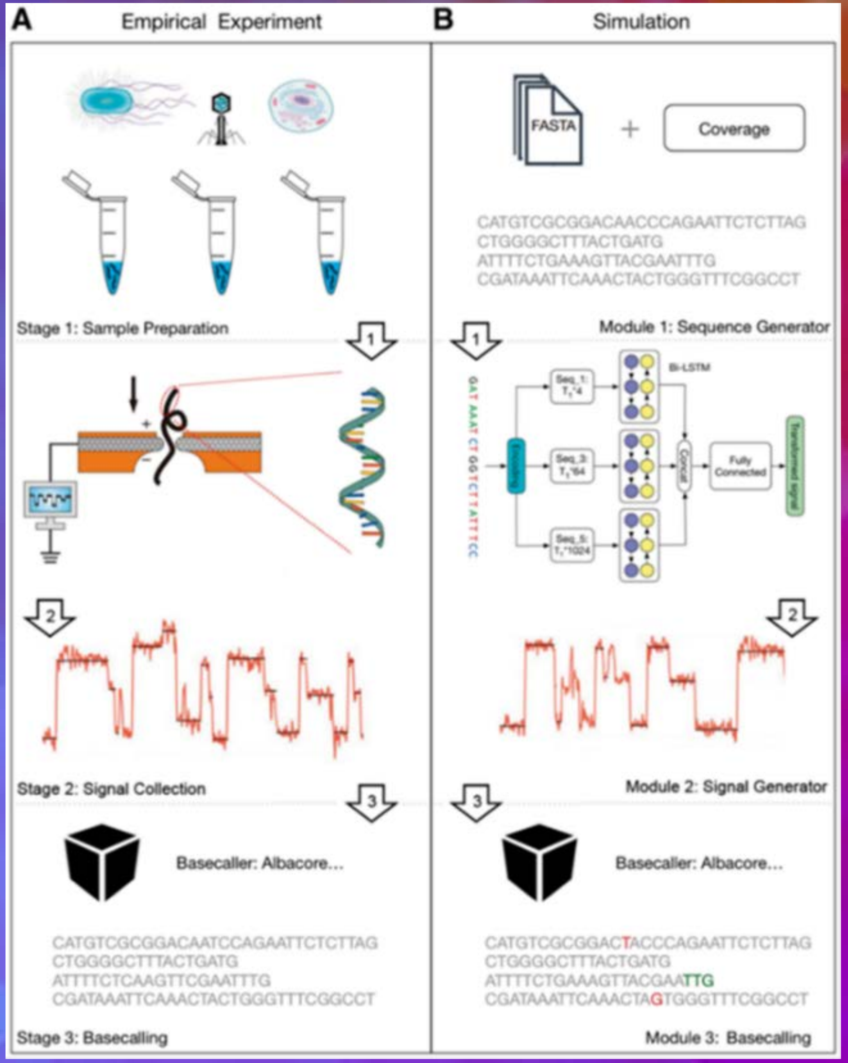
CONTEXTUAL SPEECH GENERATION

- Stephen Hawking device – effective translation of cheek movements to cursor and mouse controls
- Machine Learning, multi-context
- Customized language models
- High accuracy at predicting syllables and words
- ACAT (Assistive Contextually Aware Toolkit) by Intel Labs
- Added Speech Synthesis

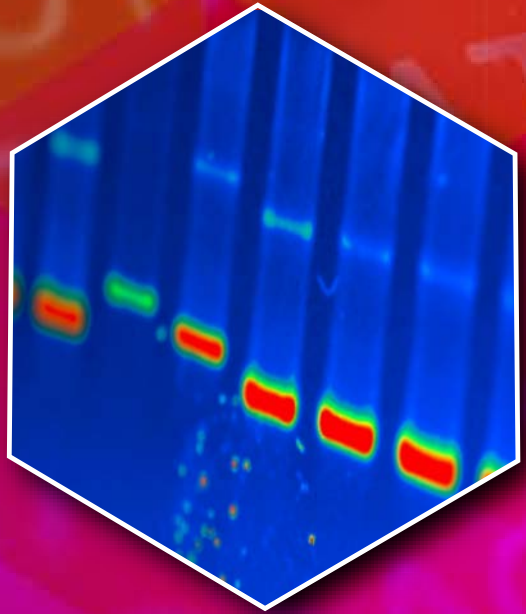
INTEL OPEN SOURCE TECHNOLOGY CENTER

https://01.org/acat

SEQUENCE MAPPING
GENOMICS - NANOPORE SEQUENCING



Yu Li et al.: *DeepSimulator: a deep simulator for Nanopore sequencing.*

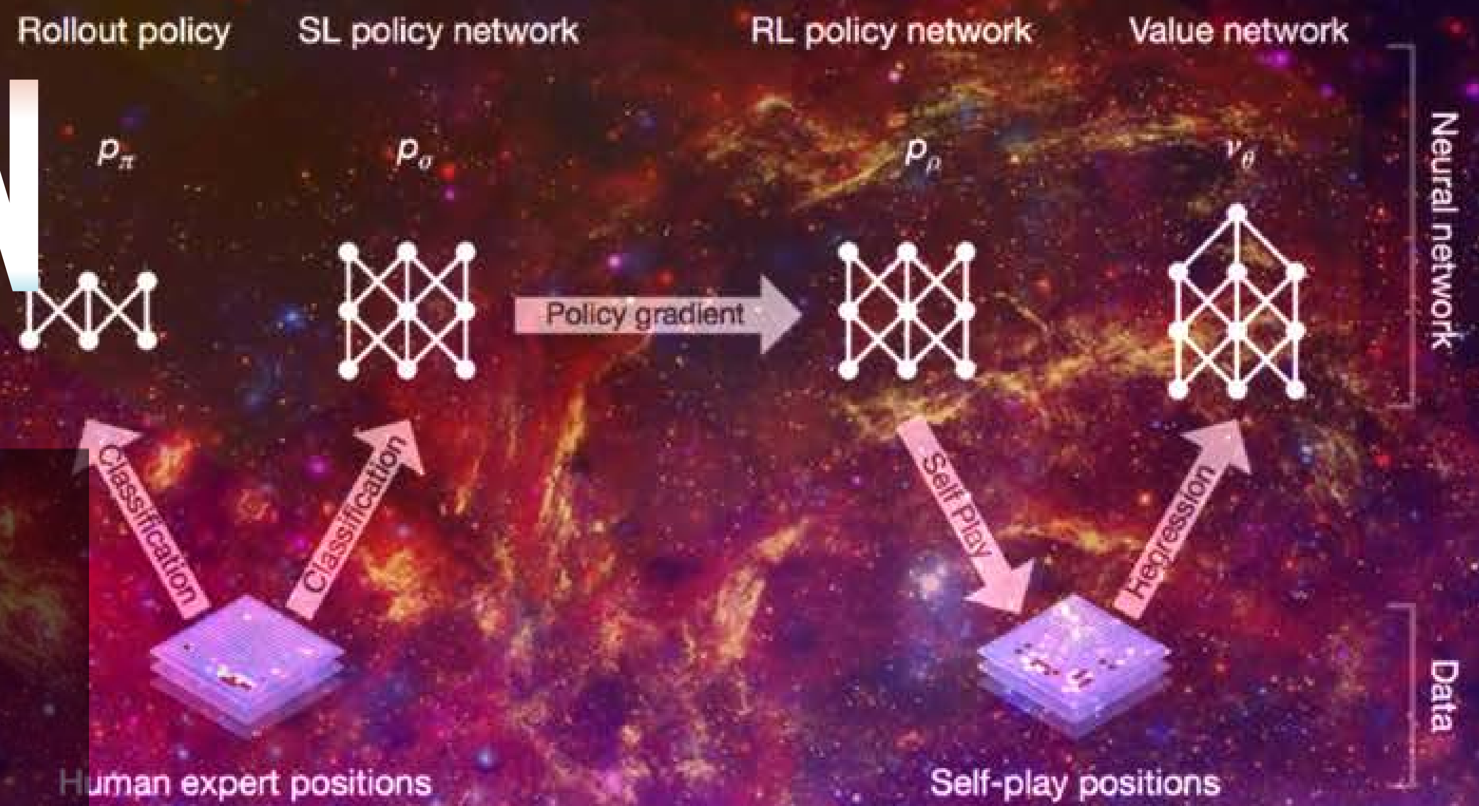


Oxford NANOPORE Technologies

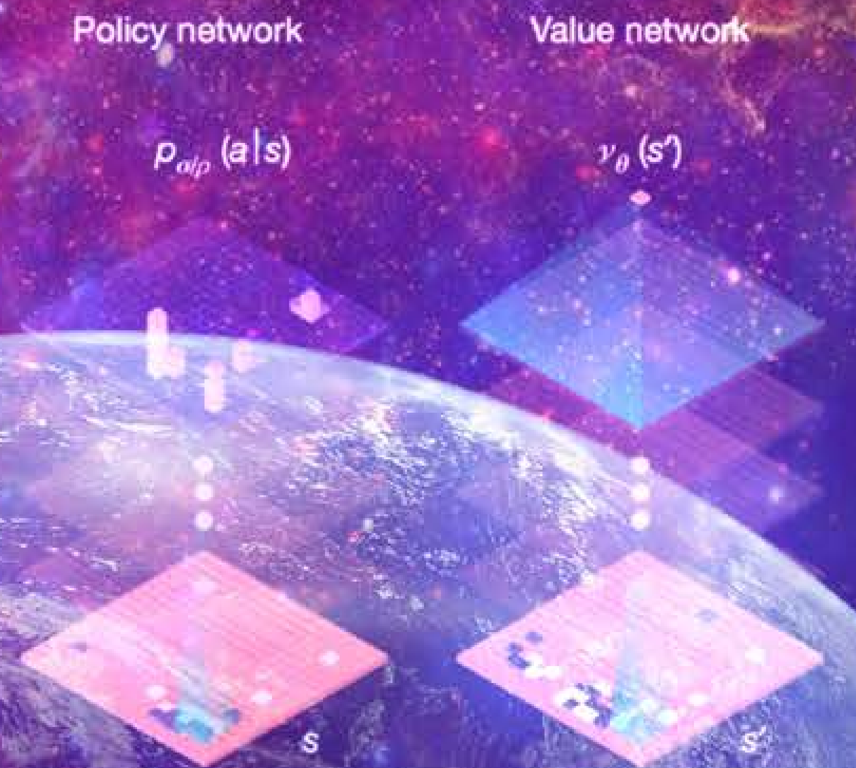
جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

- DNA/RNA high TPT sequencing by Oxford Nanopore Tech
- From noisy electrical waveforms, predicting sequence of ATCGs
- DeepSimulator mimics entire pipeline, similar to experimental
- Addressing repetitive regions

SPACE EXPLORATION



Source: https://medium.com/@karpathy/alphago-in-context-c47718cb95a5



- Solution space too large for scientist trial-and-error
- Lack of model to guide exploration

Applicable DL techniques:

Reinforcement Learning (RL)

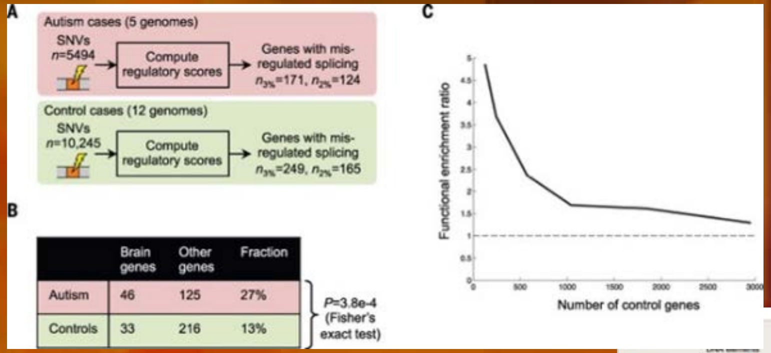
Meta Learning (learning how to best learn)

+ previous methods to evaluate branches

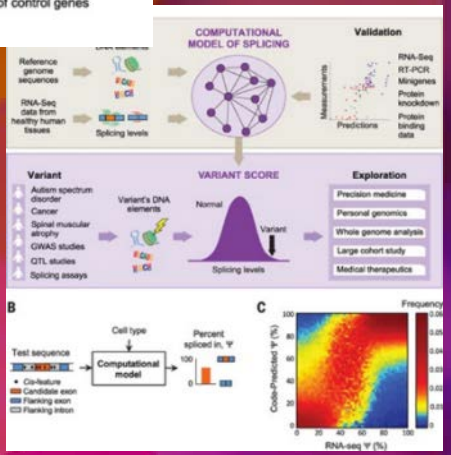
Unsupervised learning

SPACE EXPLORATION

REVEALING THE GENETIC ORIGINS OF DISEASE



Hui Y. Xiong et al. Science 2015;347:1254806

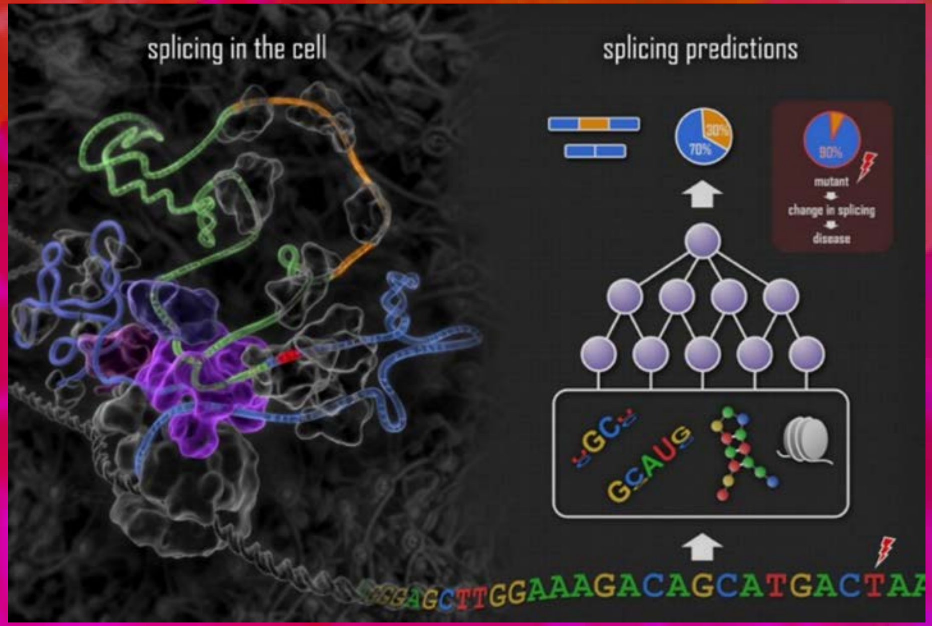


The Edward S. Rogers Sr. Department of Electrical & Computer Engineering
UNIVERSITY OF TORONTO

Science
AAAS

- Ranking of genetic mutations based on how living cells 'read' DNA
- DL learns genetic instructions for proper splicing, protein production
- Evaluate mutations and likelihood of causing disease
- Facilitate discovery of unexpected genetic determinants of autism, cancer, spinal muscular atrophy

  (H. Y. Xiong. et al.: *The human splicing code reveals new insights into genetic determinants of disease, Science 347)*
- Challenge – which mutations to try?

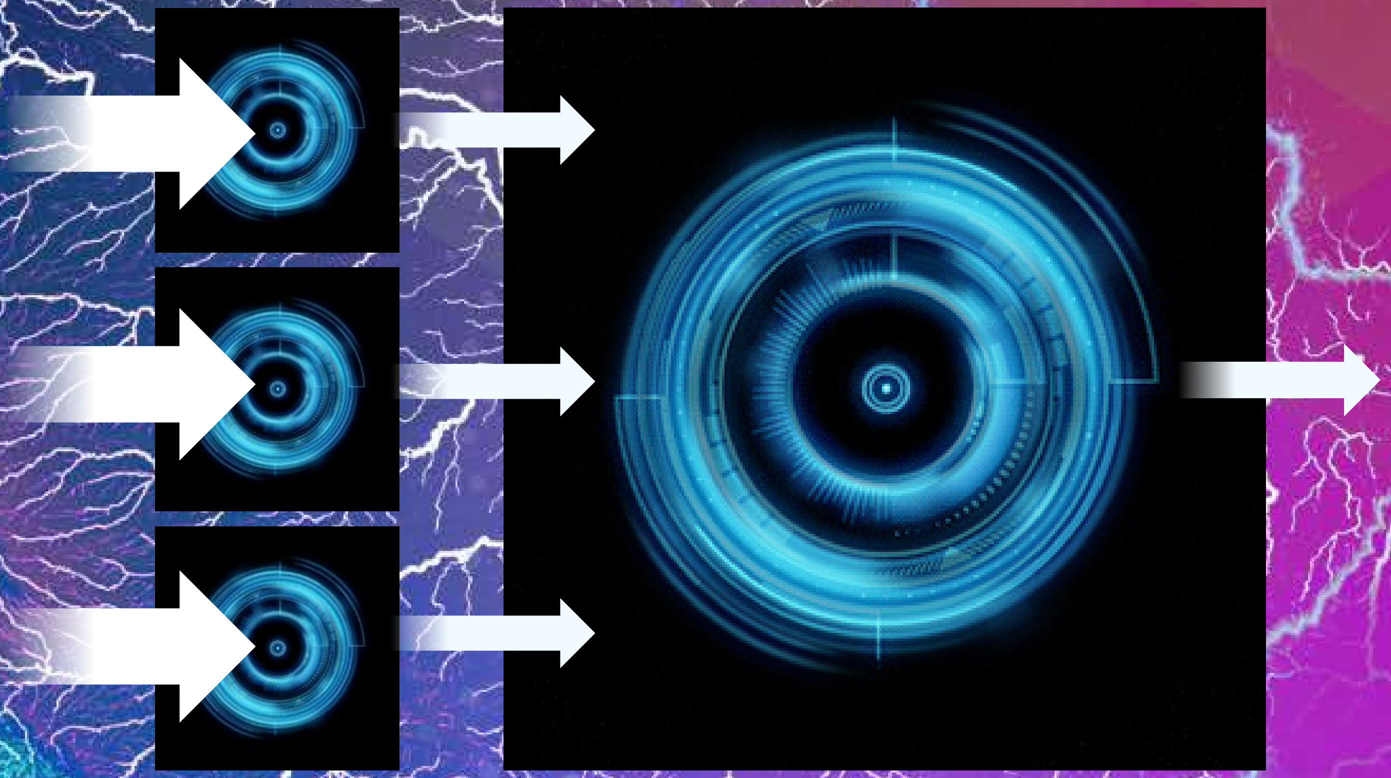


- Finding preferred path in complex space : *Neural Arch Search with Reinforcement Learning*

  (by Zoph, B. and Le, Q. V.)
- “Use ML for ML Itself”

AIDC
INTEL AI DEVCON 2018

TRIBUTARIES CURATION



- Massive number of data sources
- Data curation: intelligent filtering at the source
- Combined learning of filtering functions & data analysis

Applicable DL techniques:

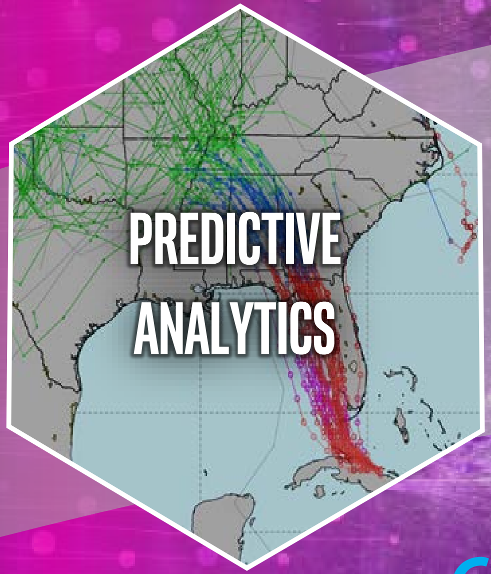
Ensemble Learning: central plus Distributed processing

Multiple ML techniques

Unsupervised Learning

FROM PAST/PRESENT TO THE FUTURE

FROM PREDICTIONS TO TAKING ACTION



ANALYTICS CURVE

PAST

PRESENT

FUTURE

OUTLOOK:
CHALLENGES AND OPPORTUNITIES

- Address lack of theory and explainability
- Understand limits of supervised & unsupervised learning
- Update skillset of senior scientists

- Fully utilize ML targeted ‘solver’ capabilities – “10^4 factor”
- Evolve from a dataset to tapping flowing phenomenon
- Harness ML as a creative co-explorer

SUMMARY
CALL TO ACTION

ANALYTICS CURVE

- DESCRIPTIVE ANALYTICS
- DIAGNOSTIC ANALYTICS
- PREDICTIVE ANALYTICS
- PRESCRIPTIVE ANALYTICS
- GENERATIVE ANALYTICS

EMBRACE AI AS A TOOL OF CHOICE AND A 'CREATIVE PARTNER'

AIDC

INTEL AI DEVCON 2018

SAN FRANCISCO | MAY 23-24