



THE AI FACTORY OF THE FUTURE

Michael Kagan, CTO | World Summit AI, 2022

INTERNET OF THINGS

20
BILLION
2021

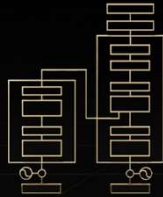


COMPUTING TECHNOLOGIES – “THE BICYCLE OF MIND”

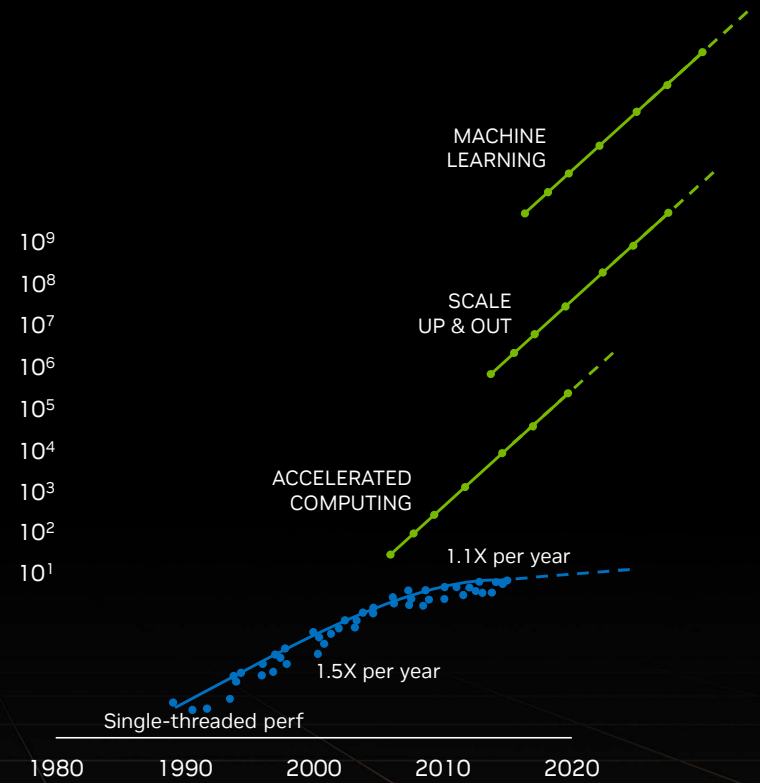
Accelerated Computing



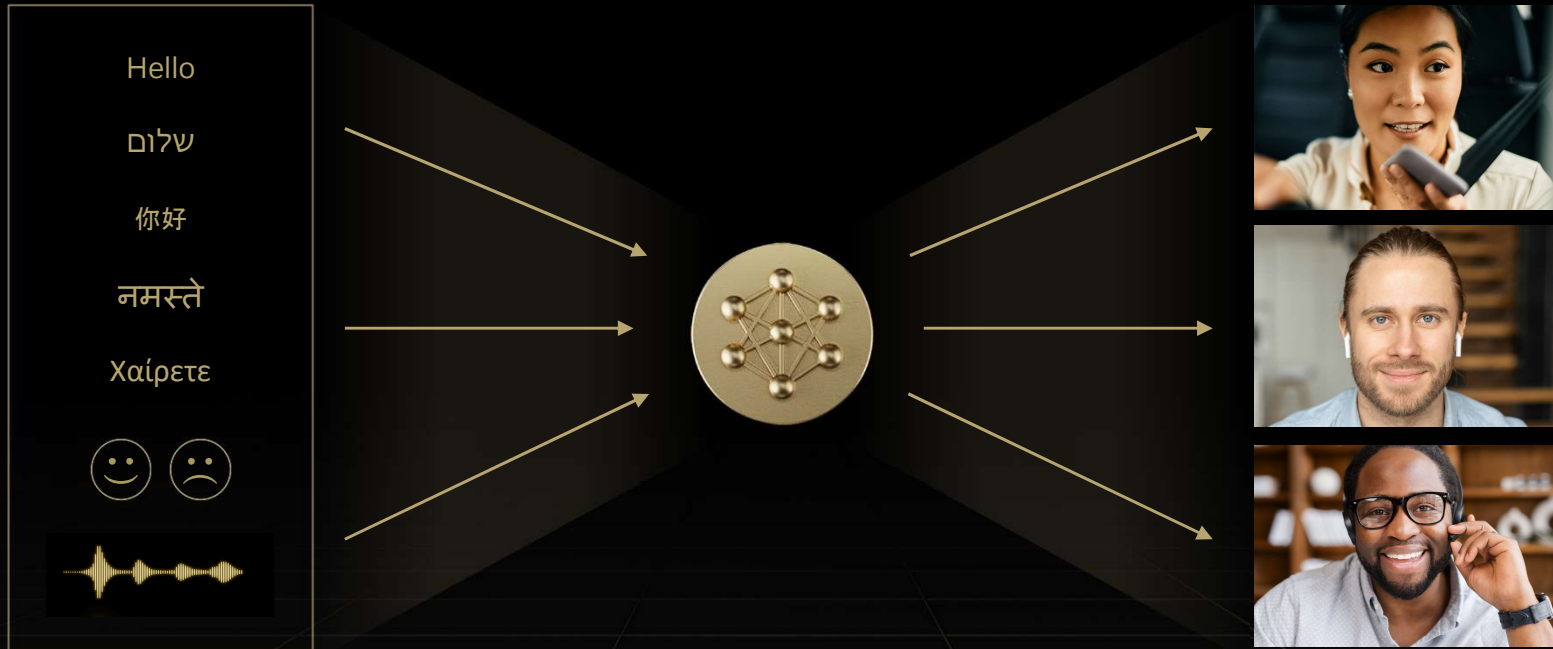
AI



Data Center Scale

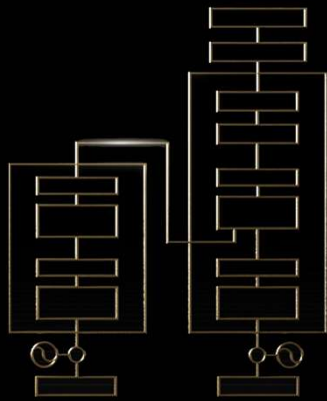


TRANSFORMERS

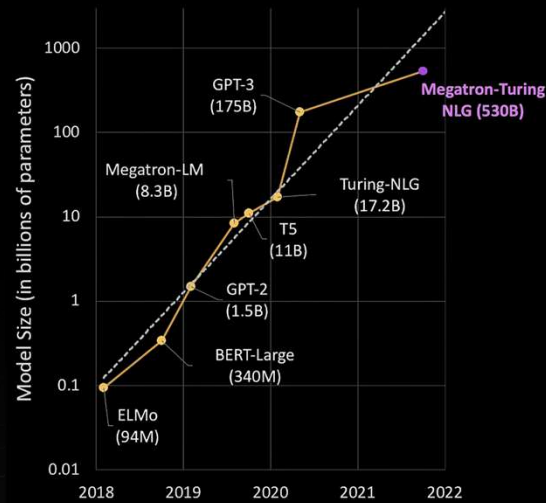


NEXT WAVE OF AI - PERFORMANCE & SCALABILITY

TRANSFORMERS TRANSFORMING AI

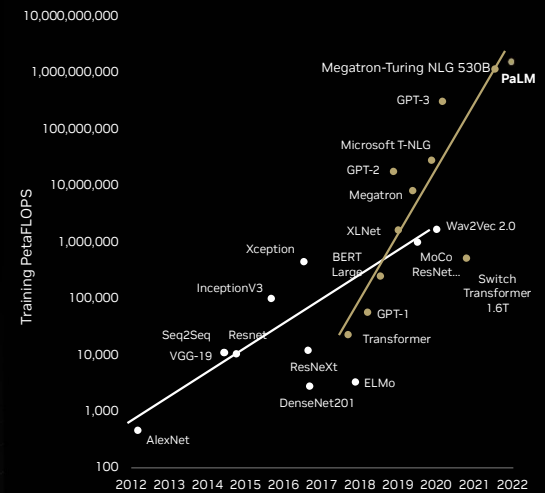


EXPLODING MODEL SIZE NEMO-Megatron Large Scale Language Models



EXPLODING COMPUTATIONAL REQUIREMENTS

Transformer AI Models = 150x / 2yrs
AI Models Excluding Transformers = 8x / 2yrs



CYBERSECURITY IS AN ENABLER

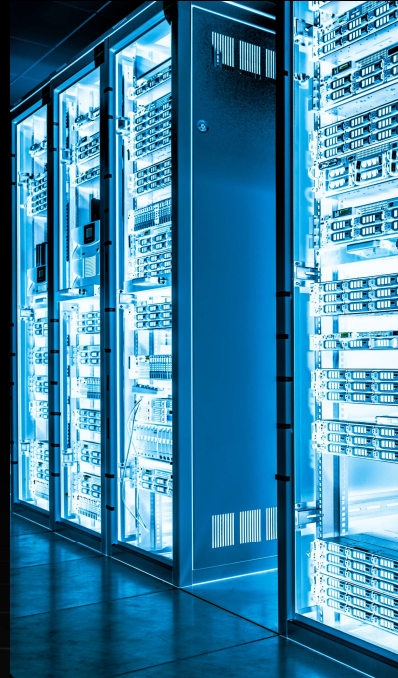
HIGH VELOCITY
DATA STREAMS



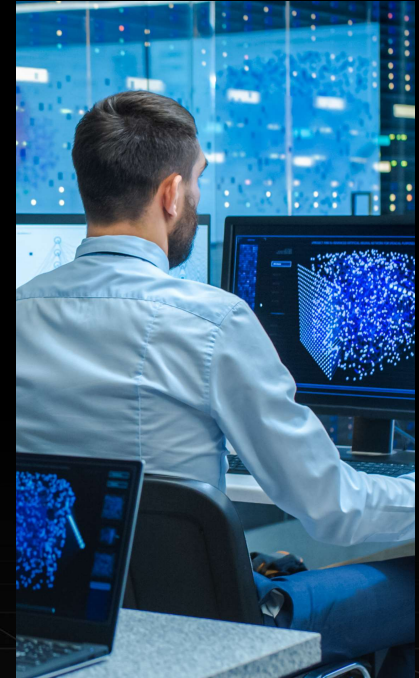
HETEROGENEOUS
DATA



CROSS ENVIRONMENT
INTEGRATIONS



QUICK ITERATIONS
AND PIVOTING

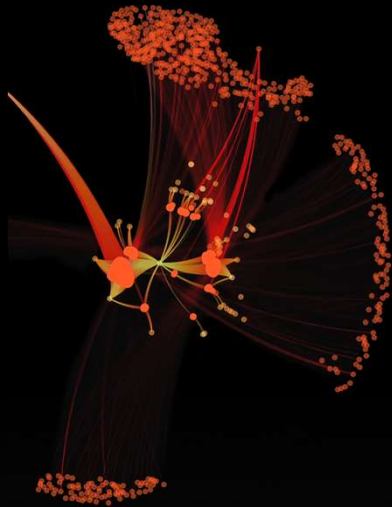


287 DAYS TO IDENTIFY A BREACH | \$4.24M AVERAGE COST PER BREACH

HUMAN-MACHINE TEAM

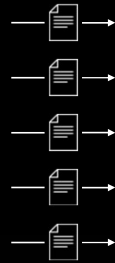
10s of Millions of Unique Models

HUMANS AND MACHINES
ACROSS THE ENTERPRISE



DIGITAL
FINGERPRINTS

LOG DATA



SECURITY ALERT
DASHBOARD

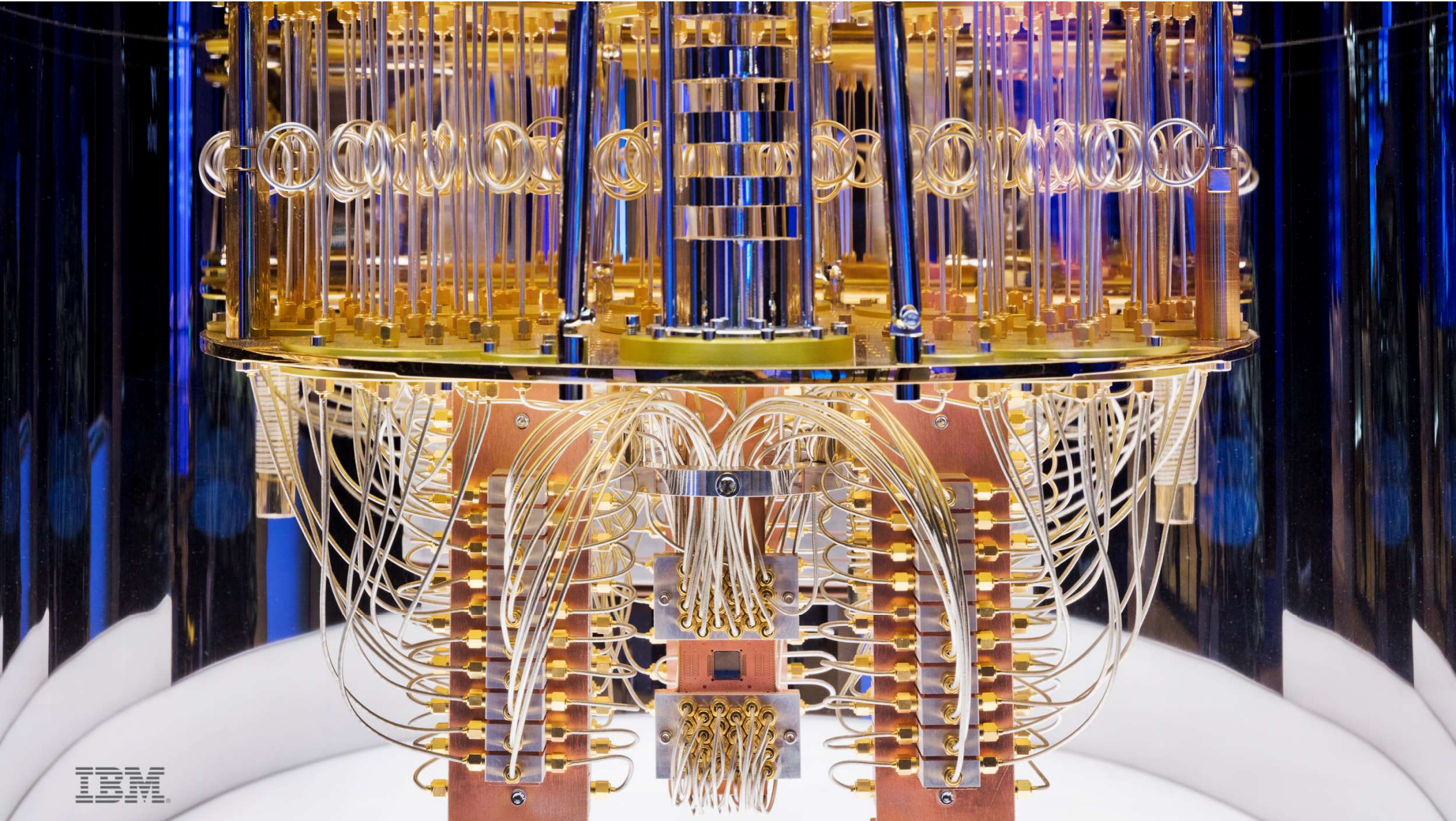
✓		Machine
✓		Human
✗		Machine
✓		Machine
✓		Machine
✗		Human
✓		Machine



ANALYZE BEHAVIOR OF EVERY HUMAN AND MACHINE
10s of Millions of Unique Models

ANOMALY DETECTION



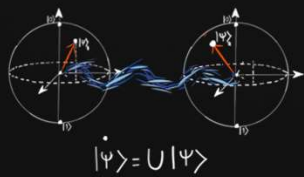


IBM

CUQUANTUM

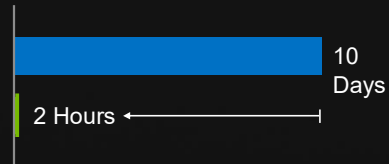
Research the Computer of Tomorrow on the Most Powerful Computer Today

GPU-ACCELERATED QUANTUM SIMULATIONS



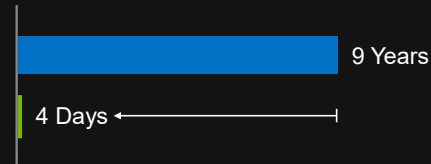
STATE VECTOR SIMULATION

Scales to 10's of Qubits



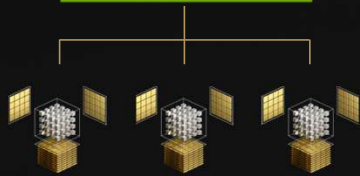
TENSOR NETWORK SIMULATION

Scales to 1000's of Qubits



■ Dual-CPU ■ DGX A100

cuQuantum



“ Using the Cotengra/Quimb packages, NVIDIA's new cuQuantum SDK, and the Selene supercomputer, we've generated a sample of the Sycamore quantum circuit at depth=20 in record time (less than 10 minutes). This sets the benchmark for quantum circuit simulation performance and will help advance the field of quantum computing by improving our ability to verify the behavior of quantum circuits. ”

—Johnnie Gray, Research Scientist, Caltech
Garnet Chan, Bren Professor of Chemistry, Caltech



Footnotes: State Vector- 1,000 circuits, 36 qubits depth =10, complex 64 | CPU: Qiskit (IBM) on Dual AMD EPYC 7742 | GPU: Qgate (NVAITC) on DGX-A100 Tensor Network - 53 qubits, depth 20 | CPU: Estimated Quimb (Caltech) on Dual AMD EPYC 7742 | GPU: Quimb (Caltech) on DGX-A100

INTERNET OF SENSES

DIGITAL TWINS



NVIDIA AI



NVIDIA OMNIVERSE



3D INTERNET



INTERNET



