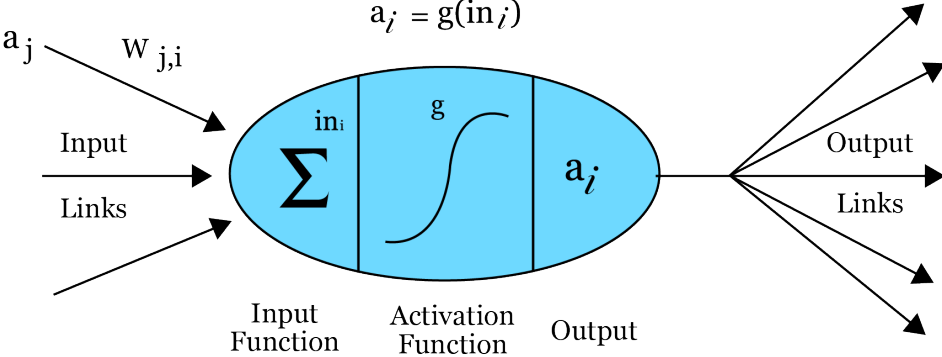
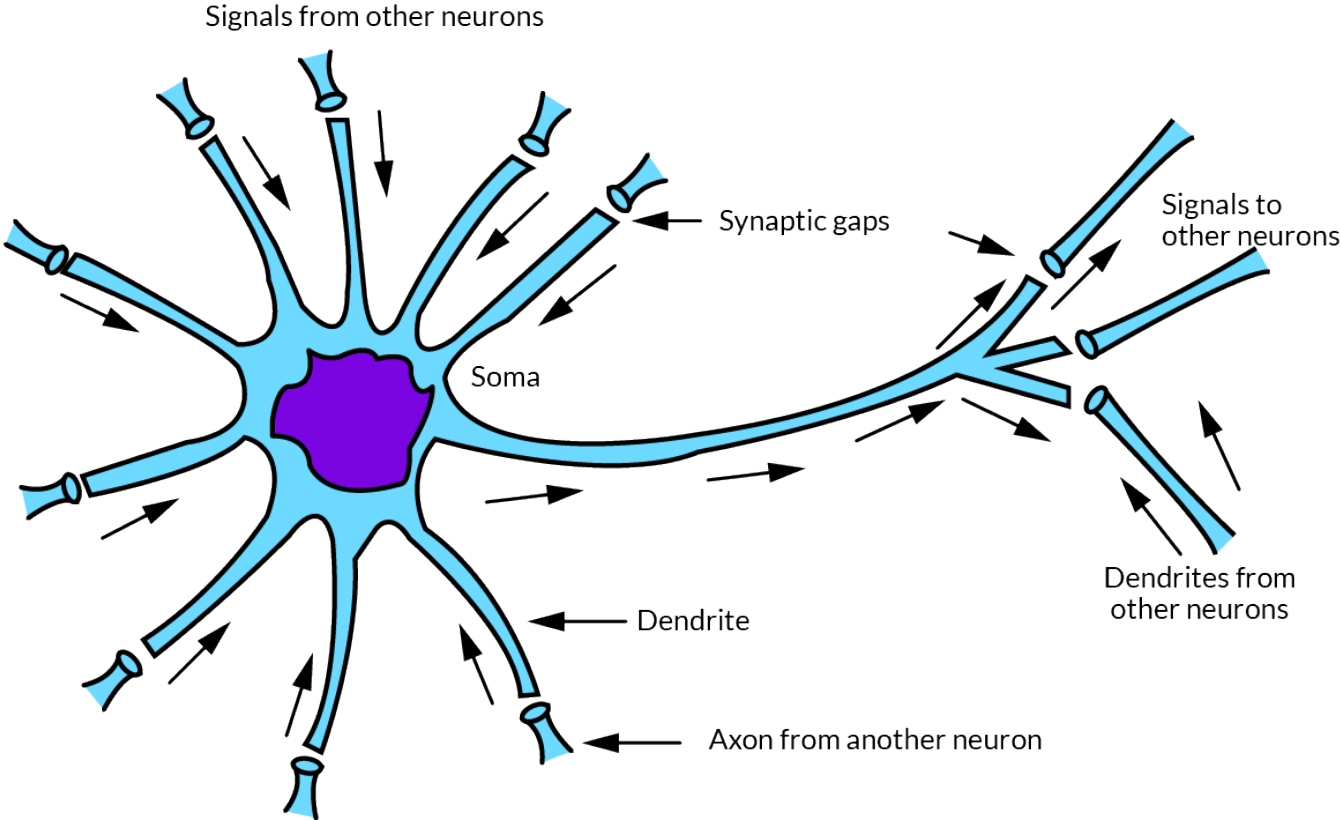


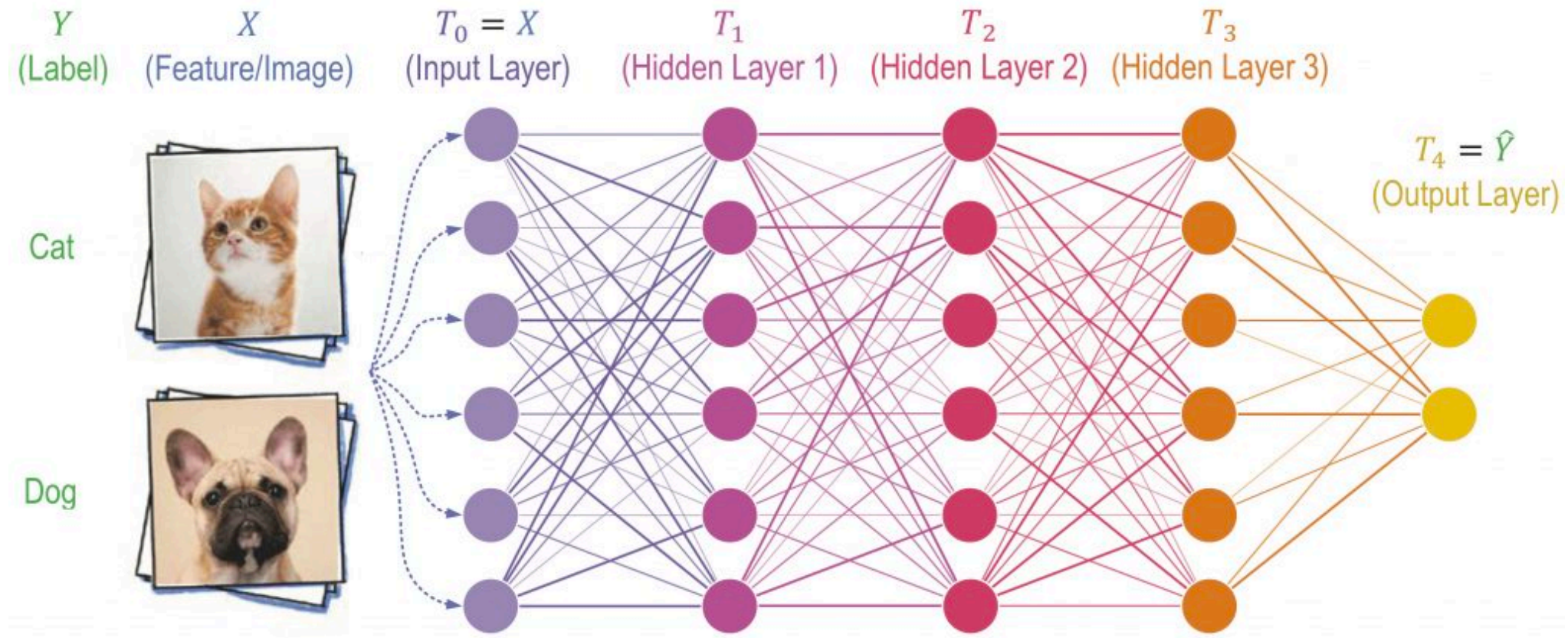
인공지능 분야
기술 중심 스타트업
비즈니스 전략

FuriosaAI
정영범 박사

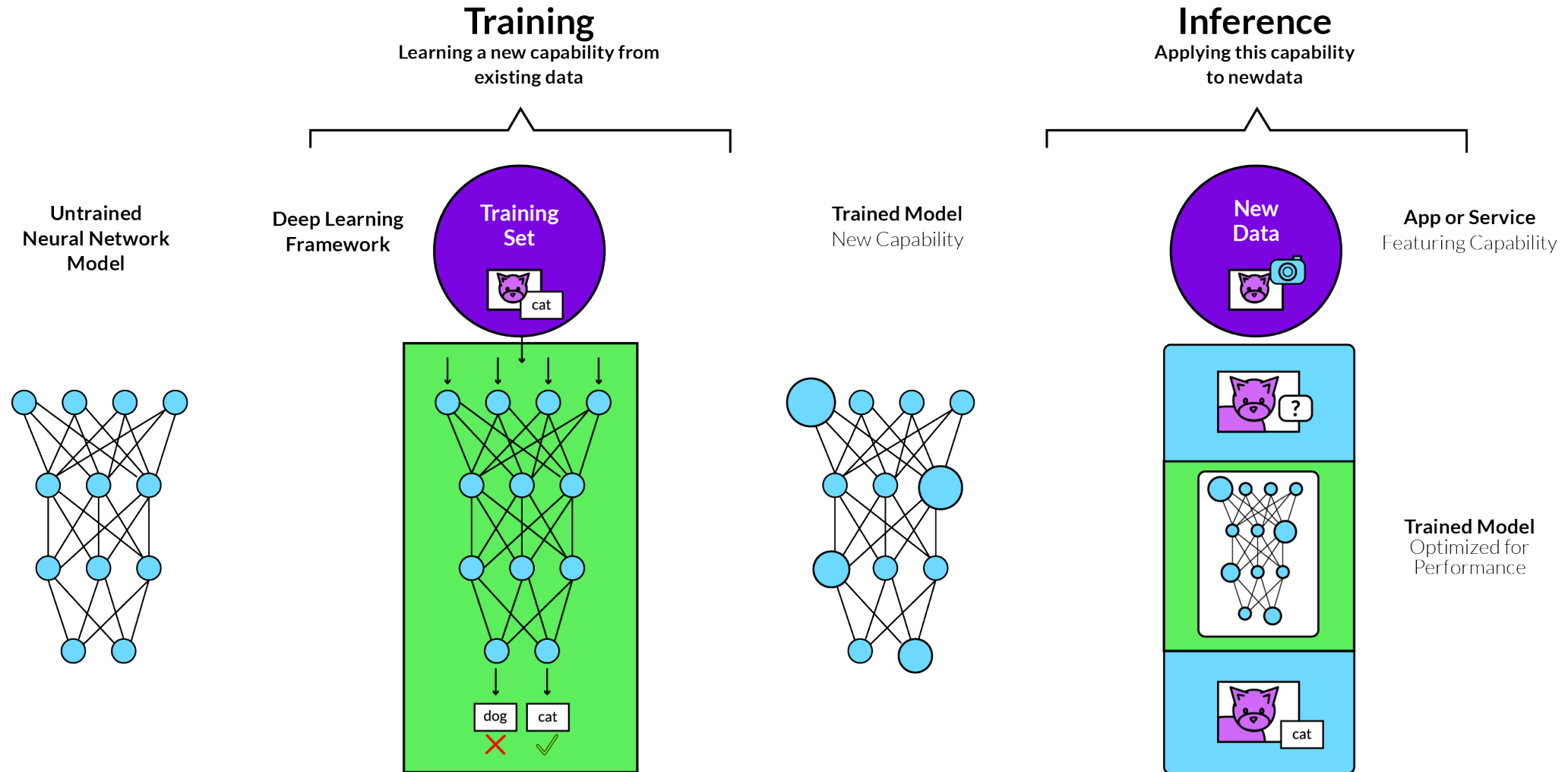
신경망 (Neural Network)



심층 신경망 (Deep Neural Network)



학습 및 추론 (Training & Inference)



AI 칩은 AI 산업의 근본적인 경쟁력을 결정 짓는 요소



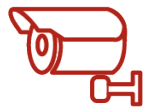
Autonomous Car



Vision Server/8k TV



Manufacturing Robot

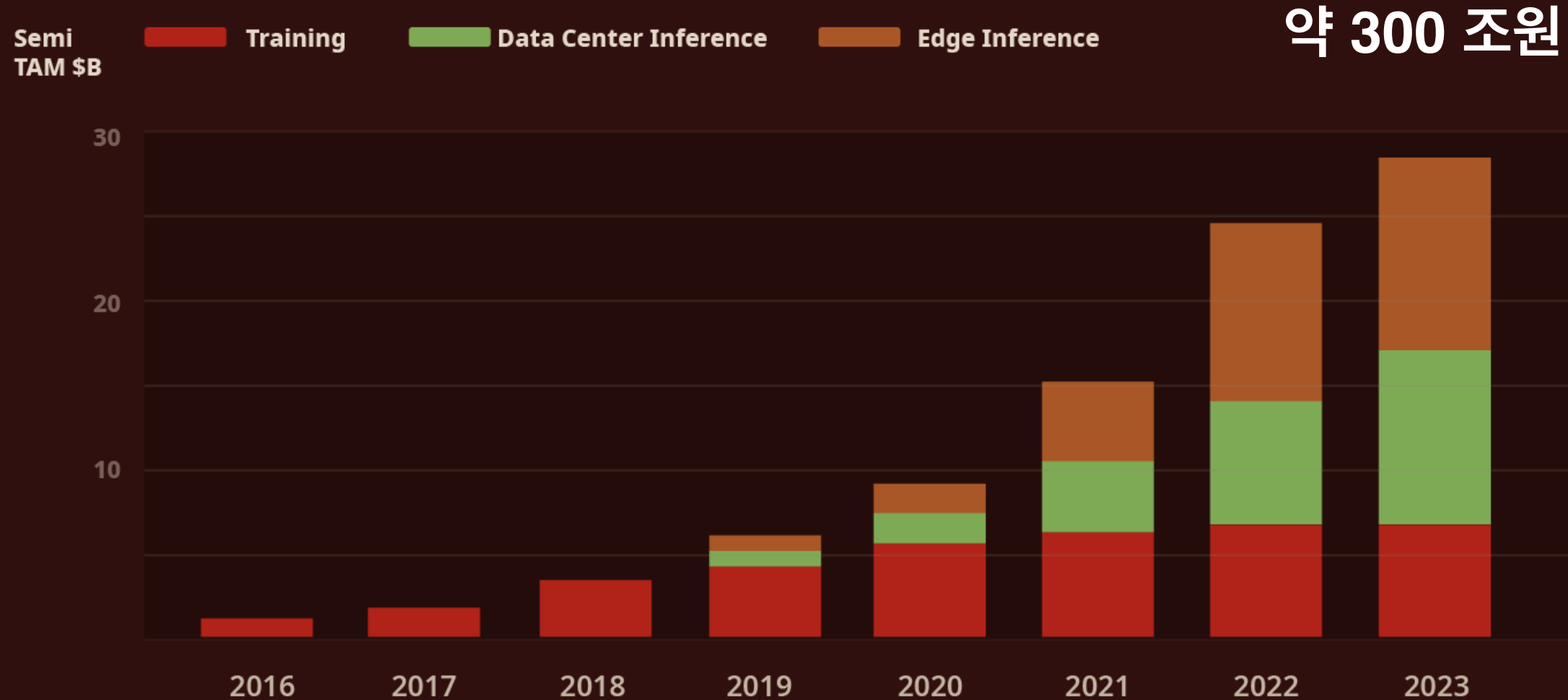


Intelligent Camera



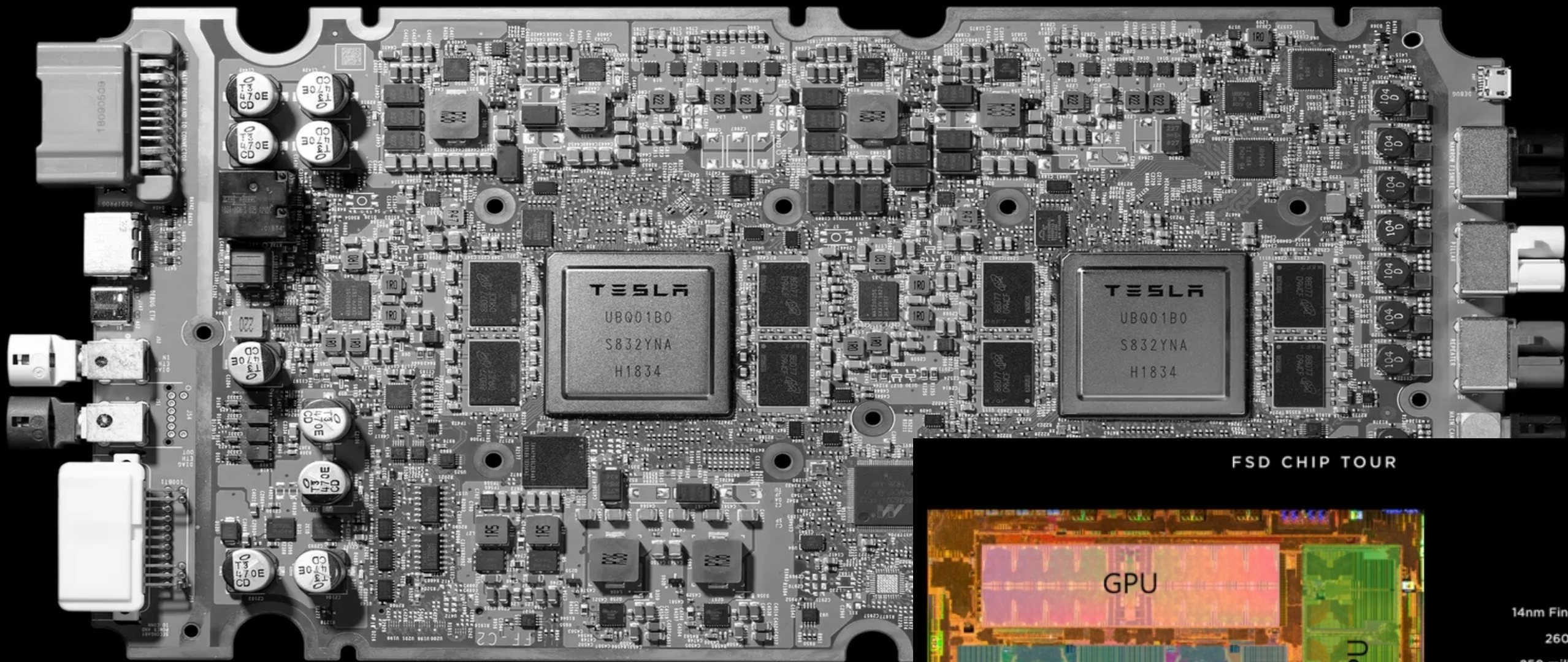
Medical Imaging Equipment

AI 칩 시장은 AI 보편화와 같이 폭발적 성장

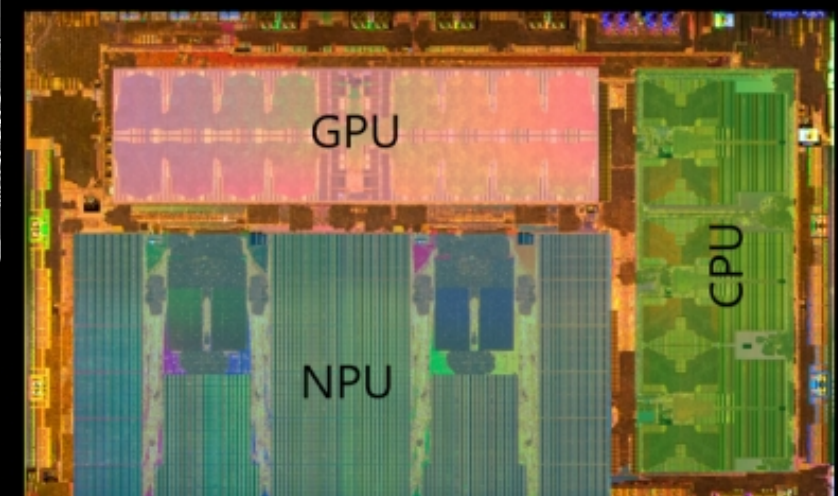


Source: Barclays Research, Company Reports May 2018

Tesla Full Self Driving Chip



FSD CHIP TOUR

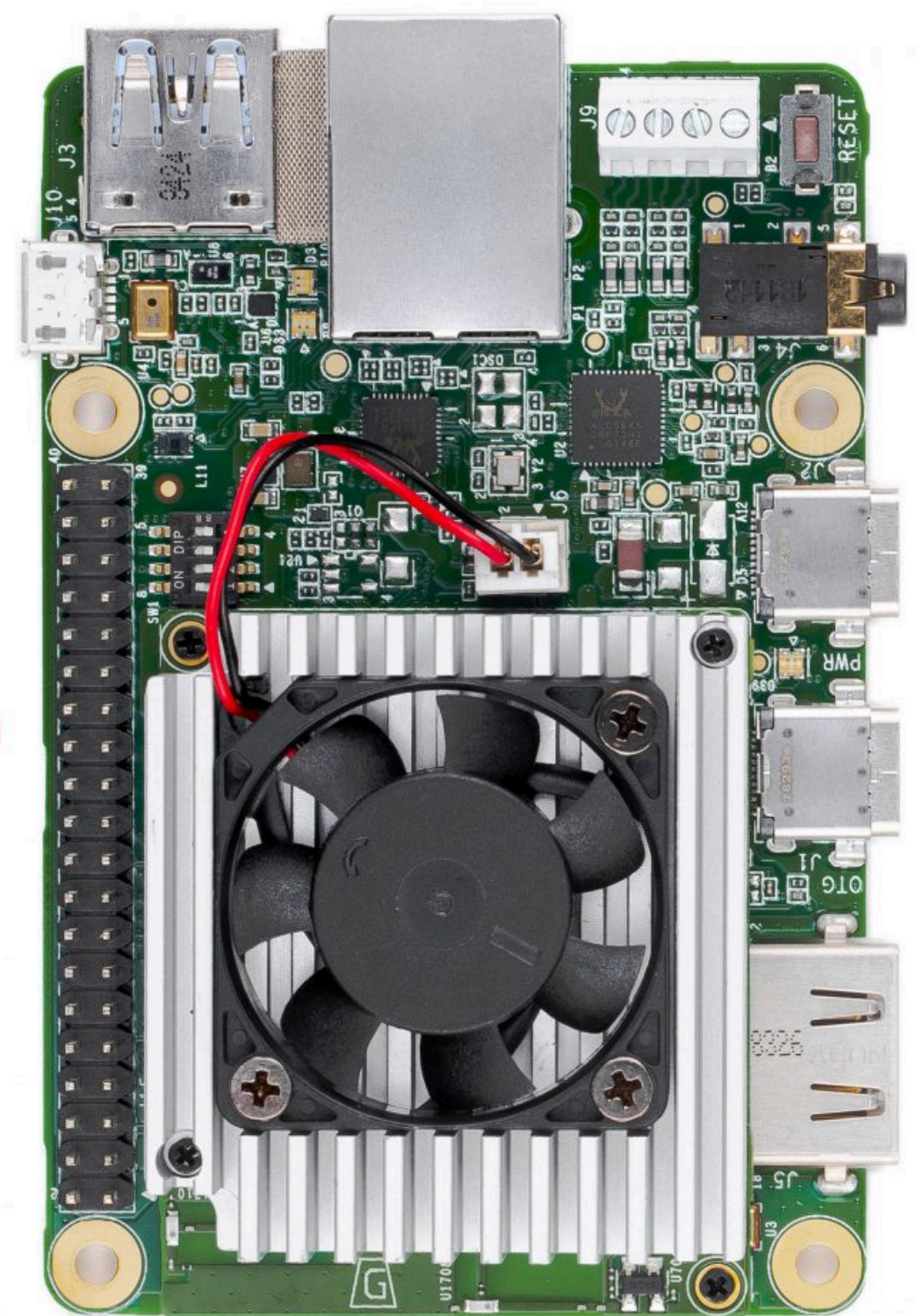


- 14nm FinFET CMOS
- 260 mm²
- 250 million gates
- 6 billion transistors
- AEC Q100

Google TPU Edge

- Edge TPU Module (SOM)
- NXP i.MX 8M SOC (Quad-core Cortex-A53, plus Cortex-M4F)
- Google Edge TPU ML accelerator coprocessor
- Vivante GC7000 GPU
- Cryptographic coprocessor
- Wi-Fi 2×2 MIMO (802.11b/g/n/ac 2.4/5GHz)
- Bluetooth 4.1
- 8GB eMMC
- 1GB LPDDR4

85 mm



56 mm



Google TPU Pod
64 2nd-gen TPUs
11.5 petaflops
4 terabytes of memory
2-D toroidal mesh network

AI 칩 시장은 글로벌 격전지 GPU 중심의 시장에 기술력 있는 회사들이 도전하는 상황

United States

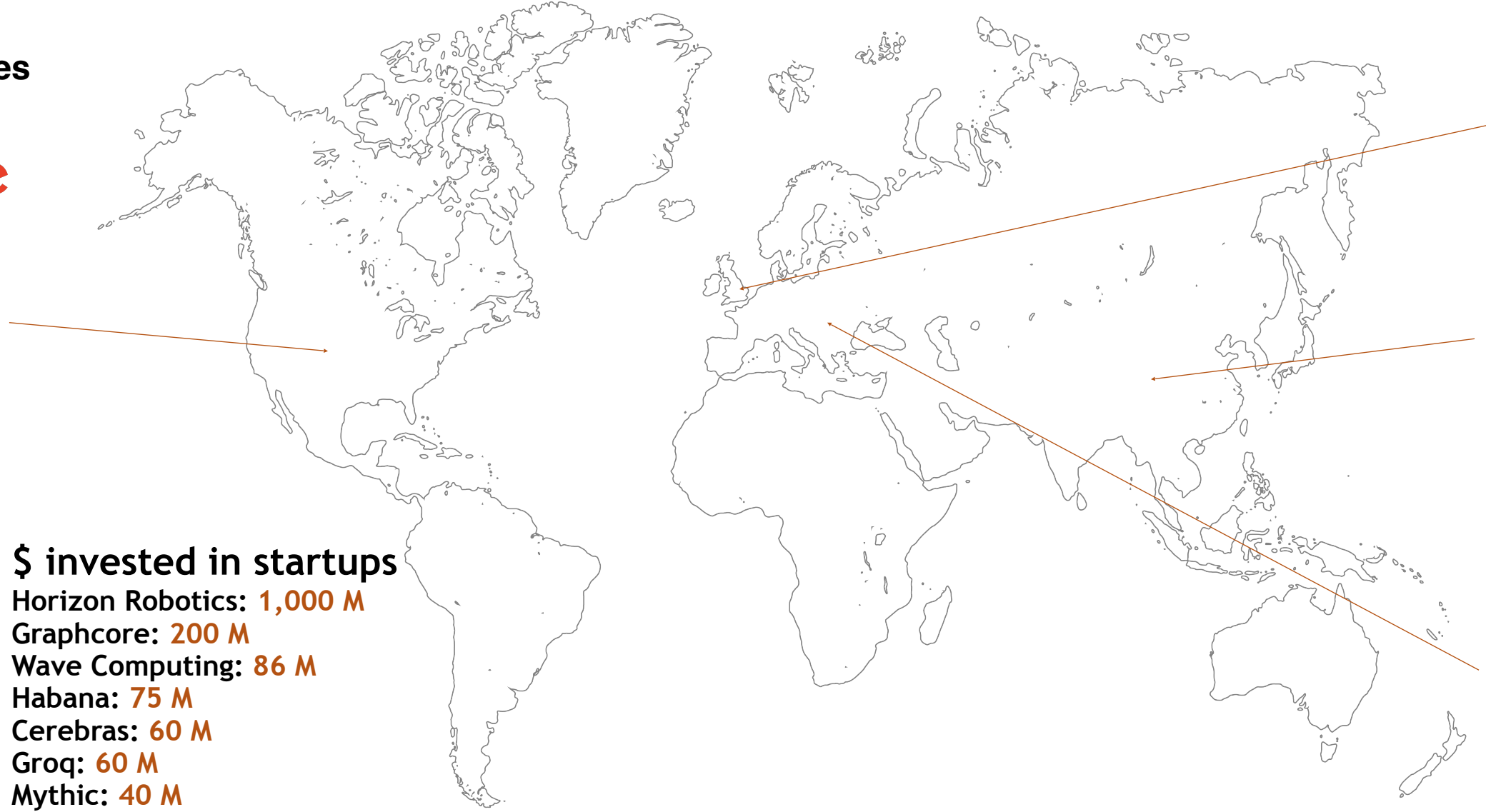


NVIDIA



\$ invested in startups

- Horizon Robotics: 1,000 M
- Graphcore: 200 M
- Wave Computing: 86 M
- Habana: 75 M
- Cerebras: 60 M
- Groq: 60 M
- Mythic: 40 M



England



China



Israel



AI 칩 개발 국내 스타트업?



백준호 CEO
Georgia Tech 석사
AMD, 삼성전자
GPU

김한준 CTO
KAIST 박사
삼성전자
Architecture

구형일
서울대 박사
아주대 교수
DNN Algorithm

2017년 4월
3명
Seed 13억 투자



글로벌 AI 칩 성능 수준 측정 MLPerf 벤치마크

2019년 7월
전 세계 26개 업체 참가 신청

2019년 10월
9개 업체 결과 제출

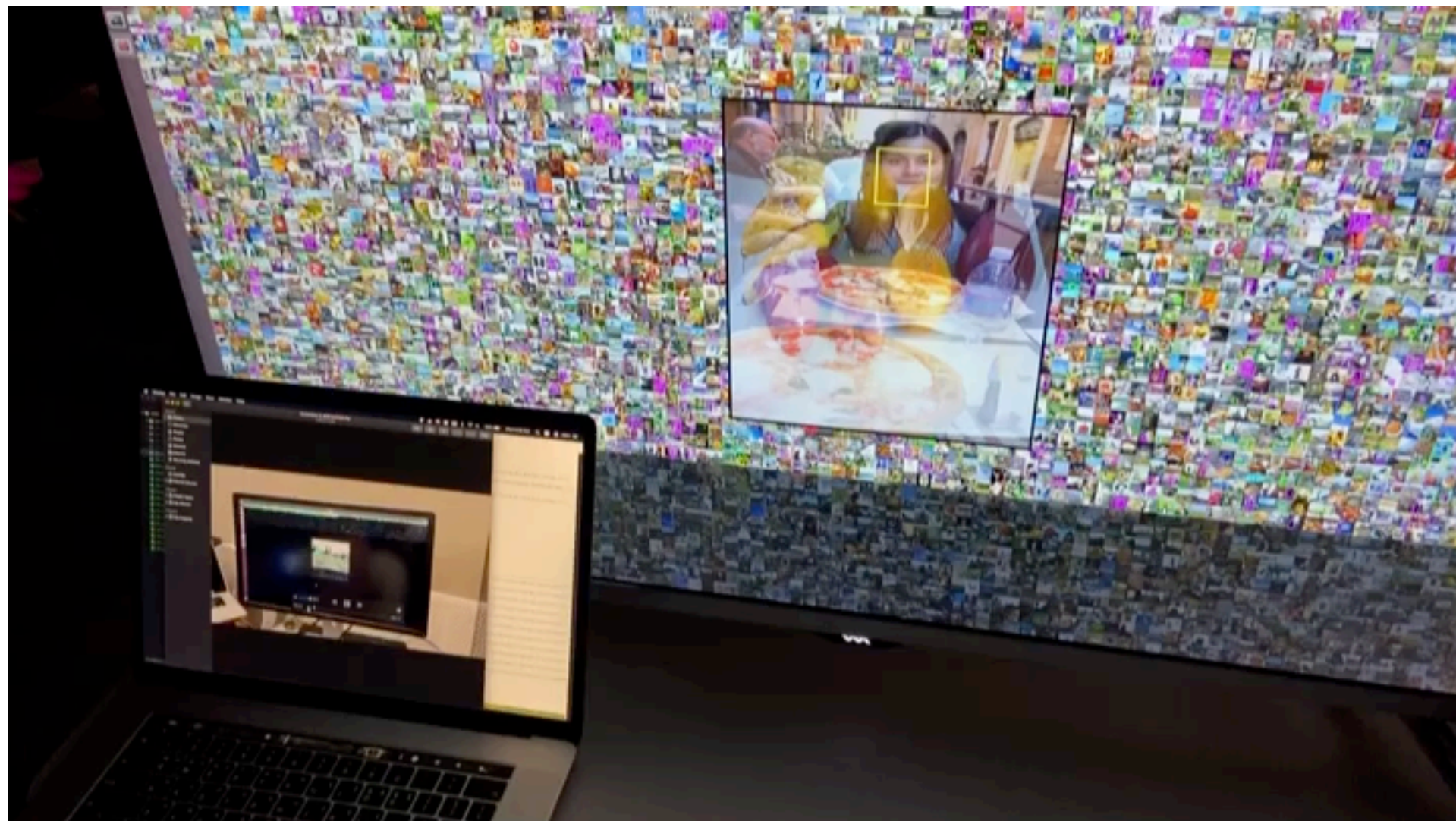


HAILO ASIC 335억 투자

FURIOSA FPGA 13억 투자

2배 이상 성능 달성

Object Detection 모델 FPGA 지원



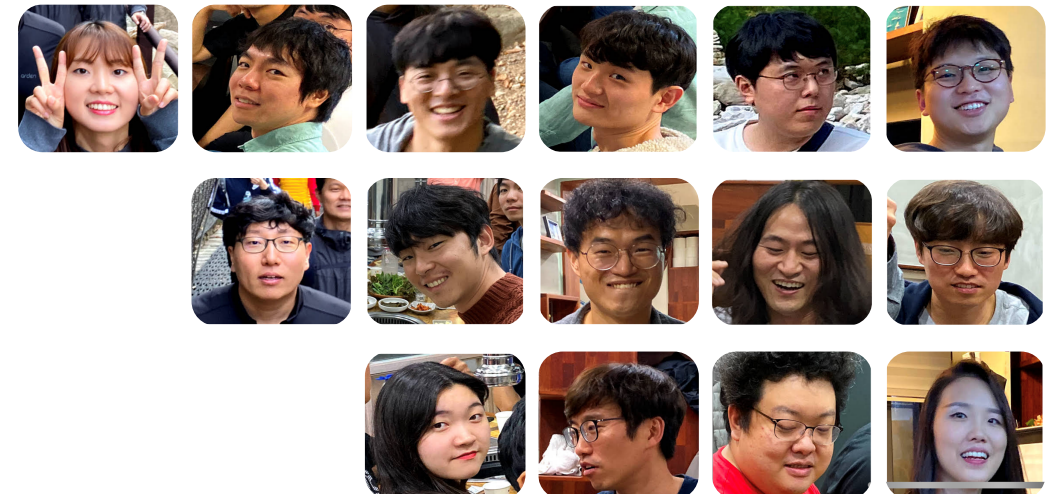
AI 칩 개발 국내 스타트업?



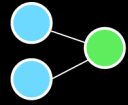
2017년 4월
3명
Seed 13억 투자



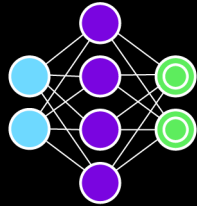
2019년 10월
25명
Series A 80억 투자



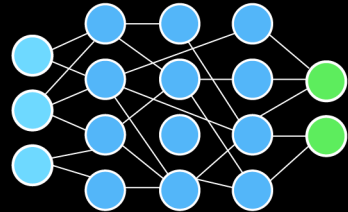
Scale of Storage



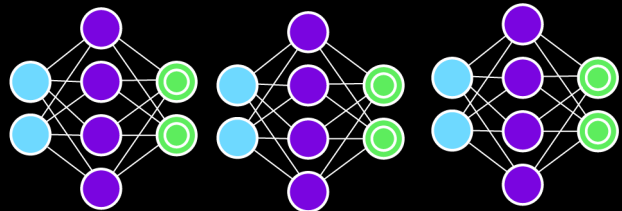
Mobile Model
> 1 MB



Speech/ Vision/Translation High Accuracy Model
> 100MB

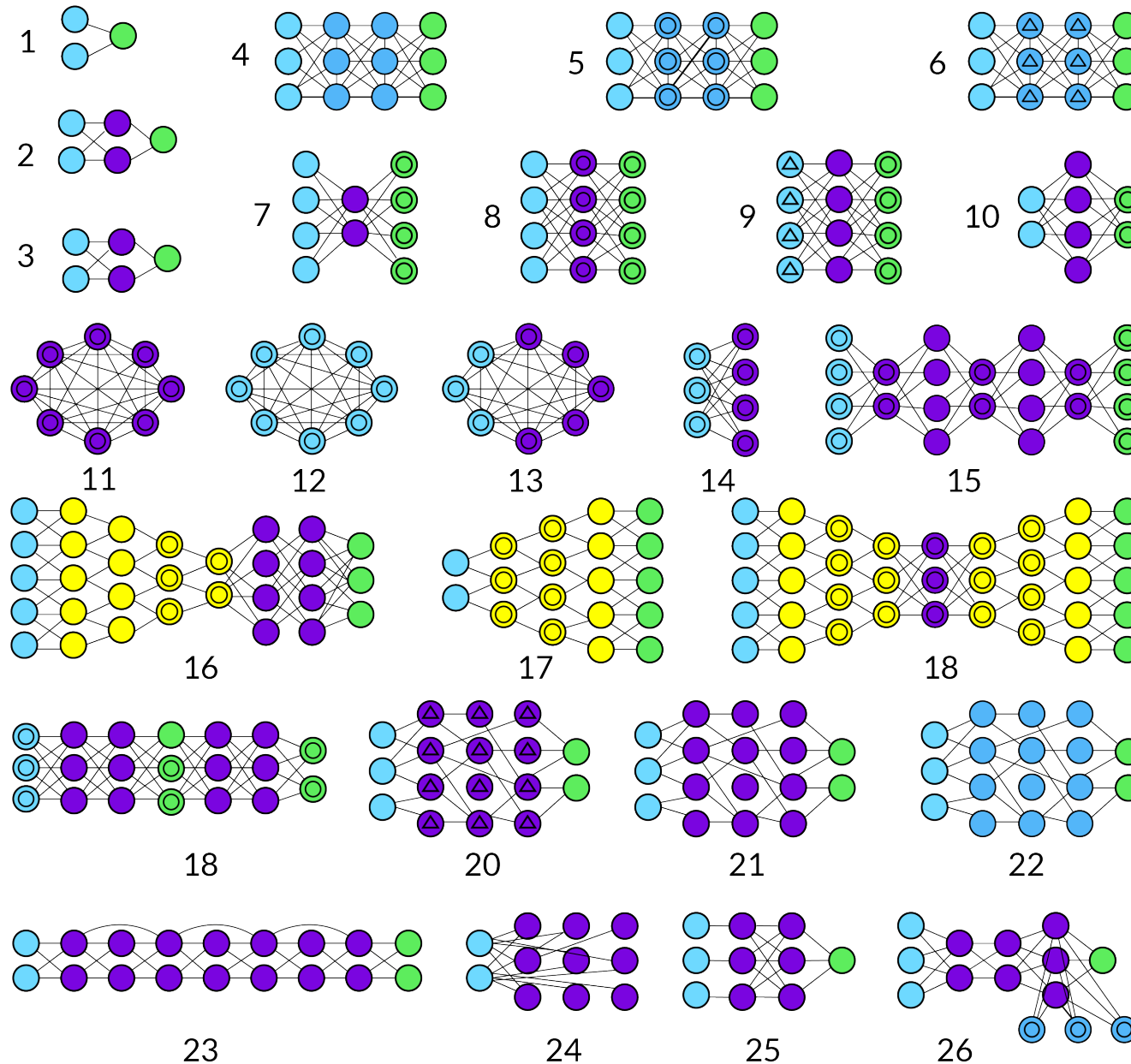


Recommendation Systems
> 1GB



Mixture of Experts
> 1TB

Scale of Model Diversities



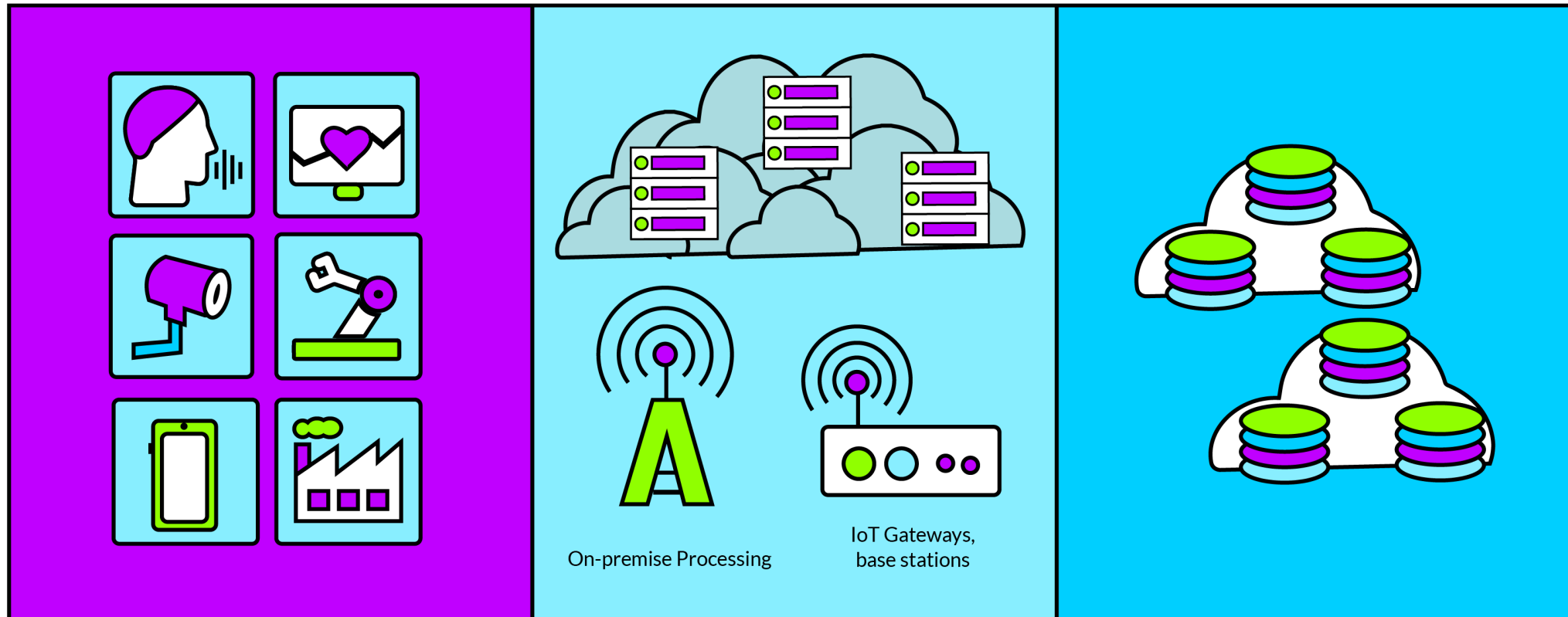
1. Perceptron (P)
2. Feed forward (FF)
3. Radial Basis Network (RBF)
4. Recurrent Neural Network (RNN)
5. Long/ Short Term Memory (LSTM)
6. Gated Recurrent Unit (GRU)
7. Auto Encoder (AE)
8. Variational AE (VAE)
9. Denoising AE (DAE)
10. Sparse AE (SAE)
11. Markov Chain (MC)
12. Hopfield Network (HN)
13. Boltzmann Machine (BM)
14. Restricted BM (RBM)
15. Deep Brief Network (DBN)
16. Deep Convolutional Network (DCN)
17. Deconvolutional Network (DN)
18. Deep Convolutional Inverse Graphics Network (DCIGN)
19. Generative Adversarial Network (GAN)
20. Liquid State Machine (LSM)
21. Extreme Learning Machine (ELM)
22. Echo state Network (ESN)
23. Deep Residual Network (DRN)
24. Kohonen Network (KN)
25. Support Vector Machine (SVM)
26. Neural Turing Machine (NTM)

Scale of Services

Connected End points

Fog / Edges

Cloud | Data Center



실시간 추론 연산을 위한 난제

Challenges

1) 대량의 성능 스케일업과 극단적인 전력 소모량 요구 사항

2) 급속도로 진화하는 AI 알고리즘

CVPR 2019 Accepts Record 1300
Papers

3) 급속도로 변하는 AI 소프트웨어 프레임워크

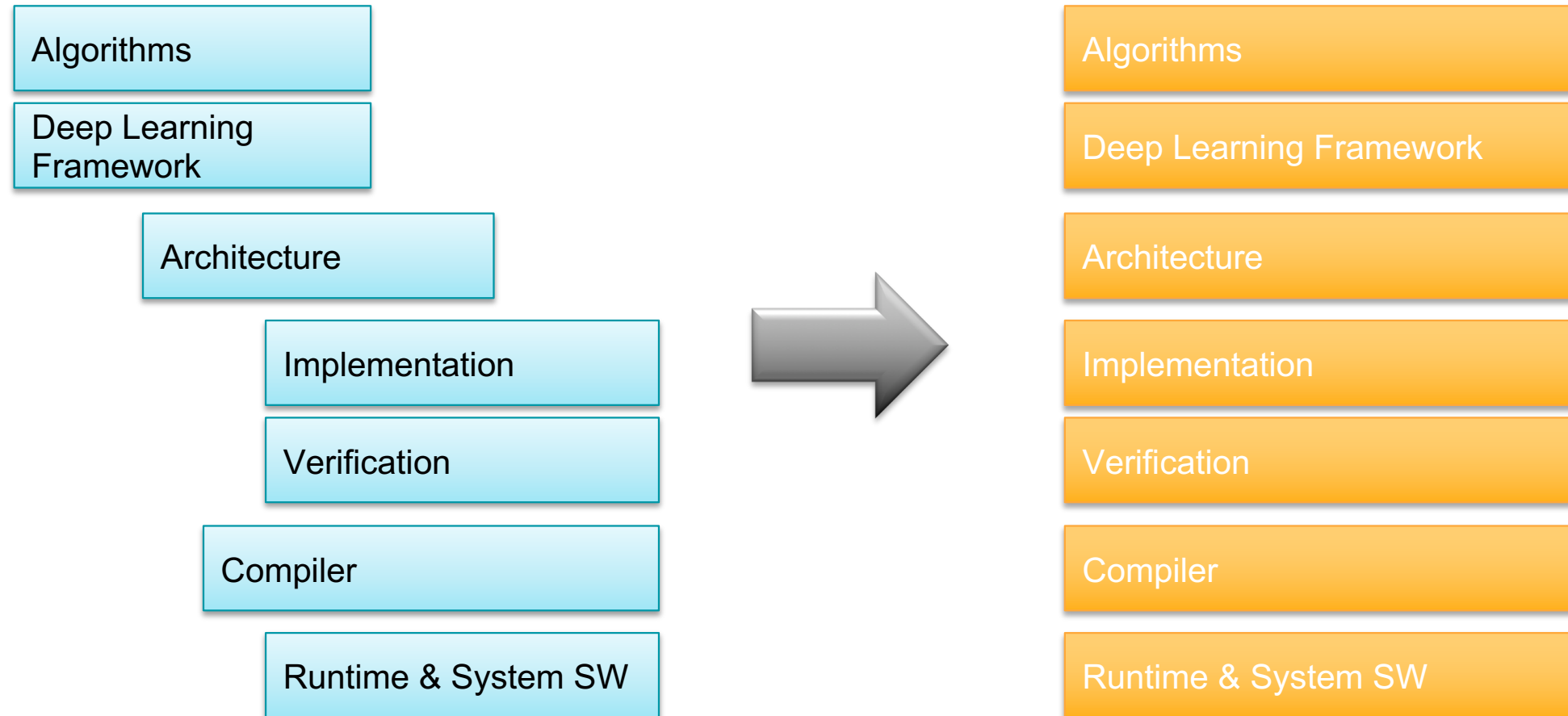
 TensorFlow

 PyTorch

Caffe

theano

AI Chip 제조 방식에 근본적인 혁신 필요



인공지능 분야 기술 중심 스타트업 비즈니스 전략

1) 최고의 기술 연구 개발에 집중

2) 최고 수준의 기술 개발을 유지할 수 있는 시스템 개발

Dream Team

3) 최고 수준의 기술 변화와 글로벌 시장 변화에 민감