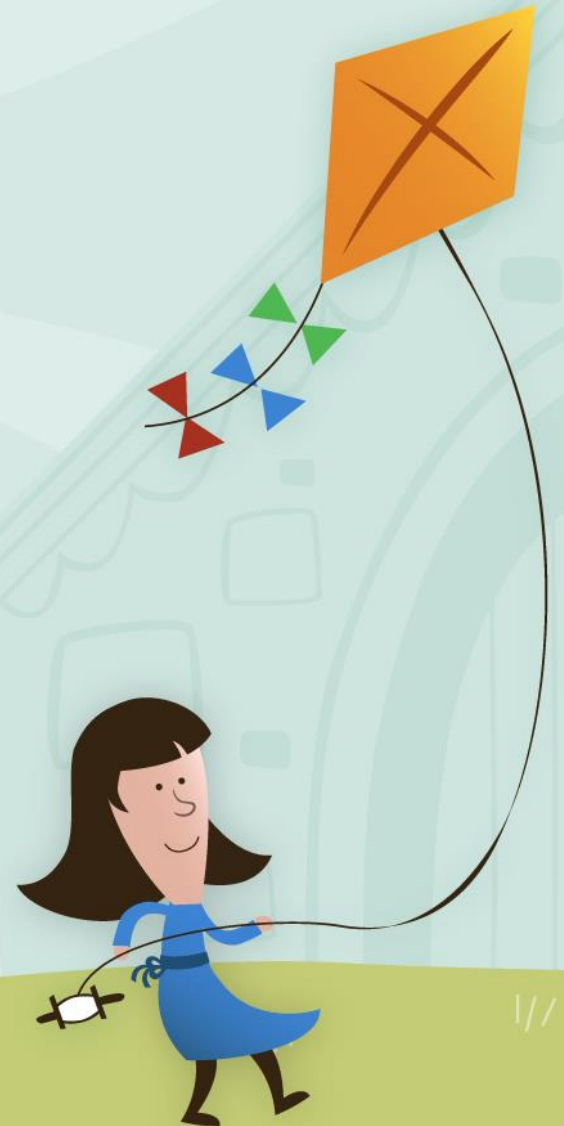


エッジ・コンピューティングの動向 ~Edge-Computing~

2019/7/25 (Thu.)



~Agenda~

1. エッジ・コンピューティングとは？
2. AIチップ
3. エッジAI製品紹介



~Summary~

1. エッジ・コンピューティングとは？

- ・クラウド側の重たい計算をエッジ側に寄せること！

2. AIチップ

- ・GAFAMは勿論、中国企業(BATIS)の台頭にも注目！

3. エッジAI製品紹介

- ・\$99で買える Jetson Nano がイチオシ！

*GAFAM(ガファム) (Google, Amazon, Facebook, Apple, Microsoft)

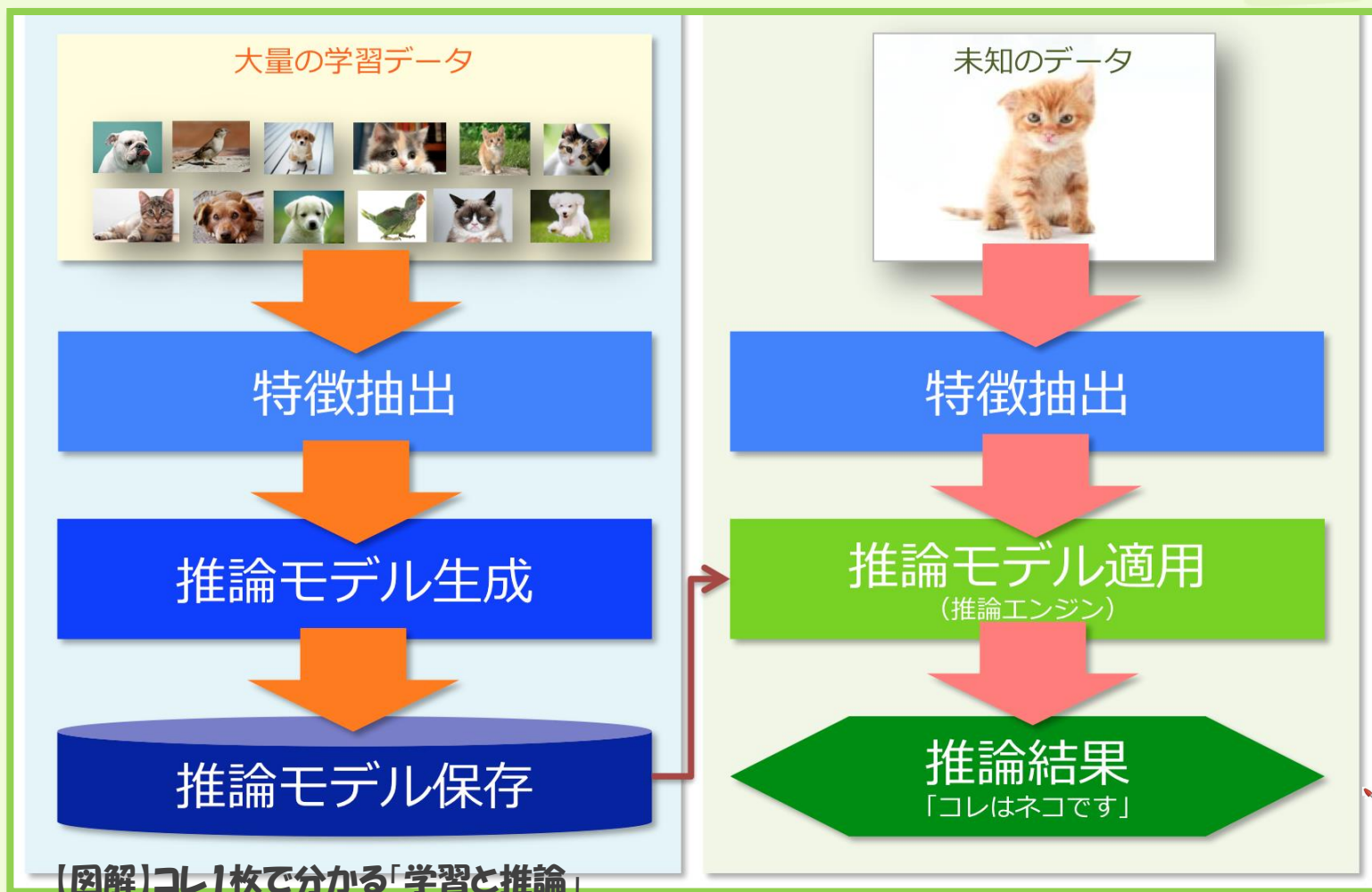
*BATIS (Baidu, Alibaba, Tencent, iFlytek, SenseTime)



【用語】 機械学習のプロセスは「学習」と「推論」から成る

学習(training)

推論(inference)



【図解】コレ1枚で分かる「学習と推論」

https://blogs.itmedia.co.jp/itsolutionjuku/2015/07/post_106.html



~Agenda~

1. エッジ・コンピューティングとは？

2. AIチップ

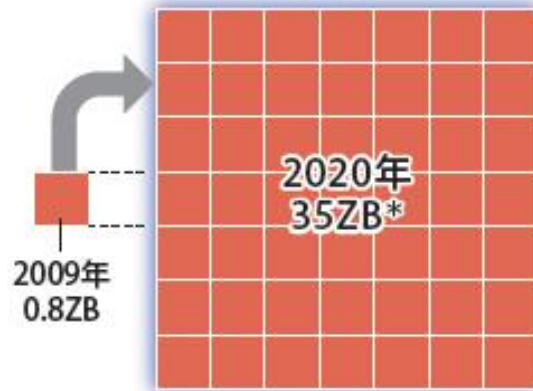
3. エッジAI製品紹介



クラウドが抱える3つの壁

① データ量の壁

約10年で
データ量が44倍!



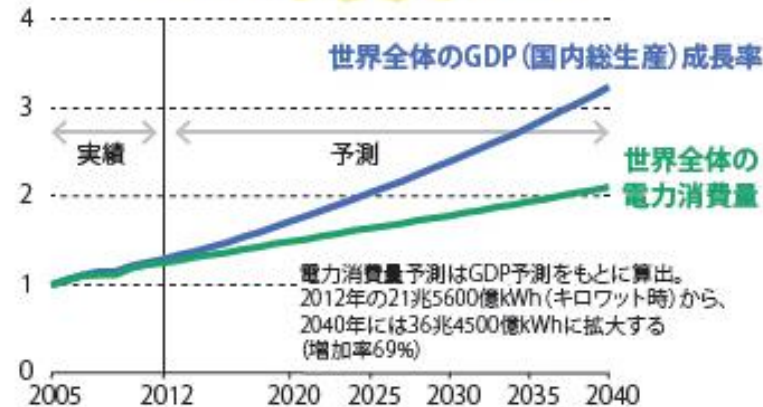
単純なIoTシステムを動かすには
ファイルサーバーの容量が足りない!

エッジ端末でデータの
取捨選択をする機能が必要!

* ZB: ゼタバイト=10の21乗バイト
出典: IDC『The Digital Universe Decade - Are You Ready?』

② エネルギーの壁

世界の電力消費量は
増え続けている!



単純なIoTシステムを動かすには
エネルギーが足りない!

エッジ端末で必要な
データを保存する機能が必要!

出典: 米Energy Information Administration
『International Energy Outlook2016』

③ リアルタイム性の壁

通信速度のボトルネックで
事故を回避できない!



単純なIoTシステムは
リアルタイム処理には対応できない!

エッジ端末で状況判断するだけの
知能を持たせることが必要!

<https://special.nikkeibp.co.jp/atcl/TEC/17p/080800023/>

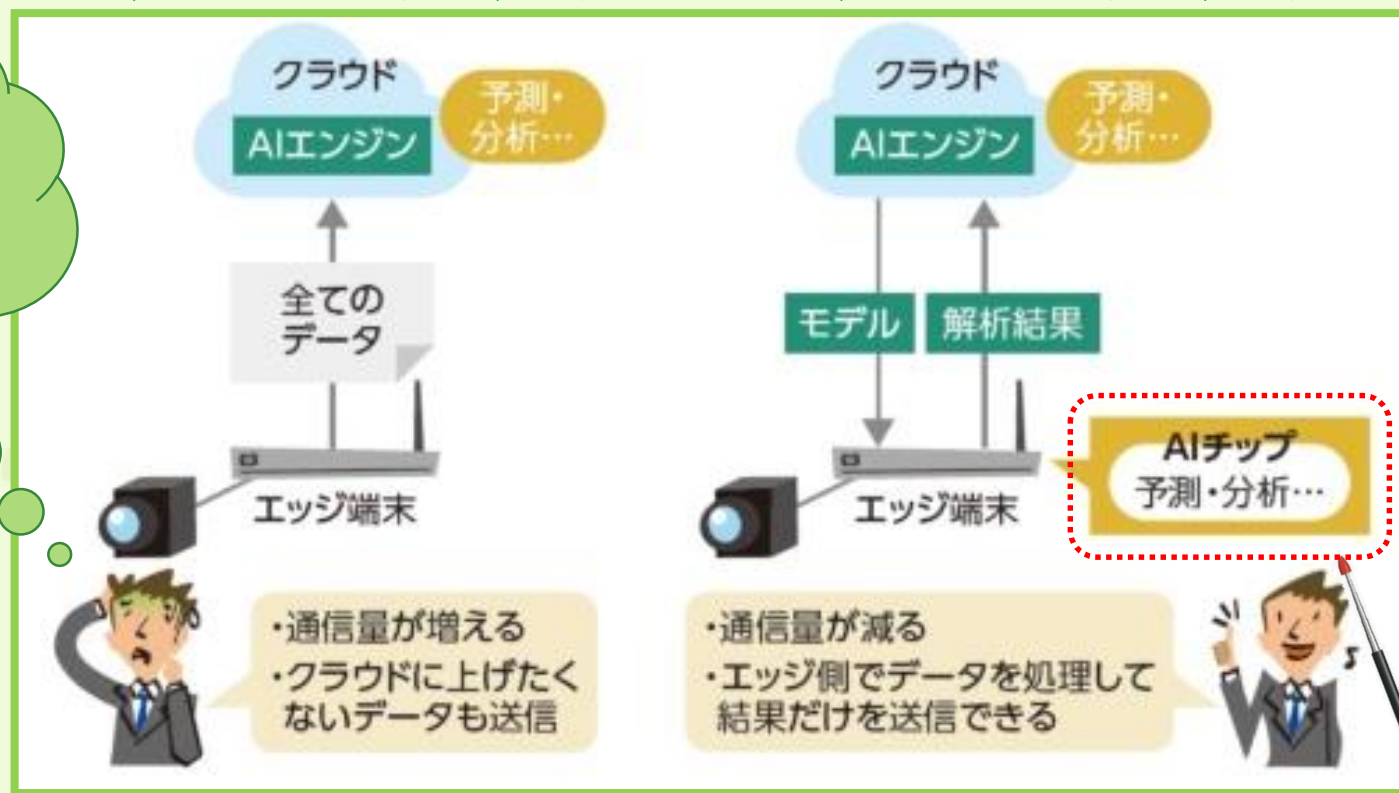


重たい計算をエッジ側に寄せてクラウドの負荷分散を図る

クラウド コンピューティング

エッジ コンピューティング

- ①データ量の壁
- ②エネルギーの壁
- ③リアルタイムの壁



注.) エッジ・コンピューティングとIoTは異なる概念！



エッジ・コンピューティングの活躍が期待されているのは エッジの近くでリアルタイムな判断が求められるケース

例1： 自動運転

- ・ 車載カメラから人を検知して、即座に自動停止

例2： 故障診断・検知

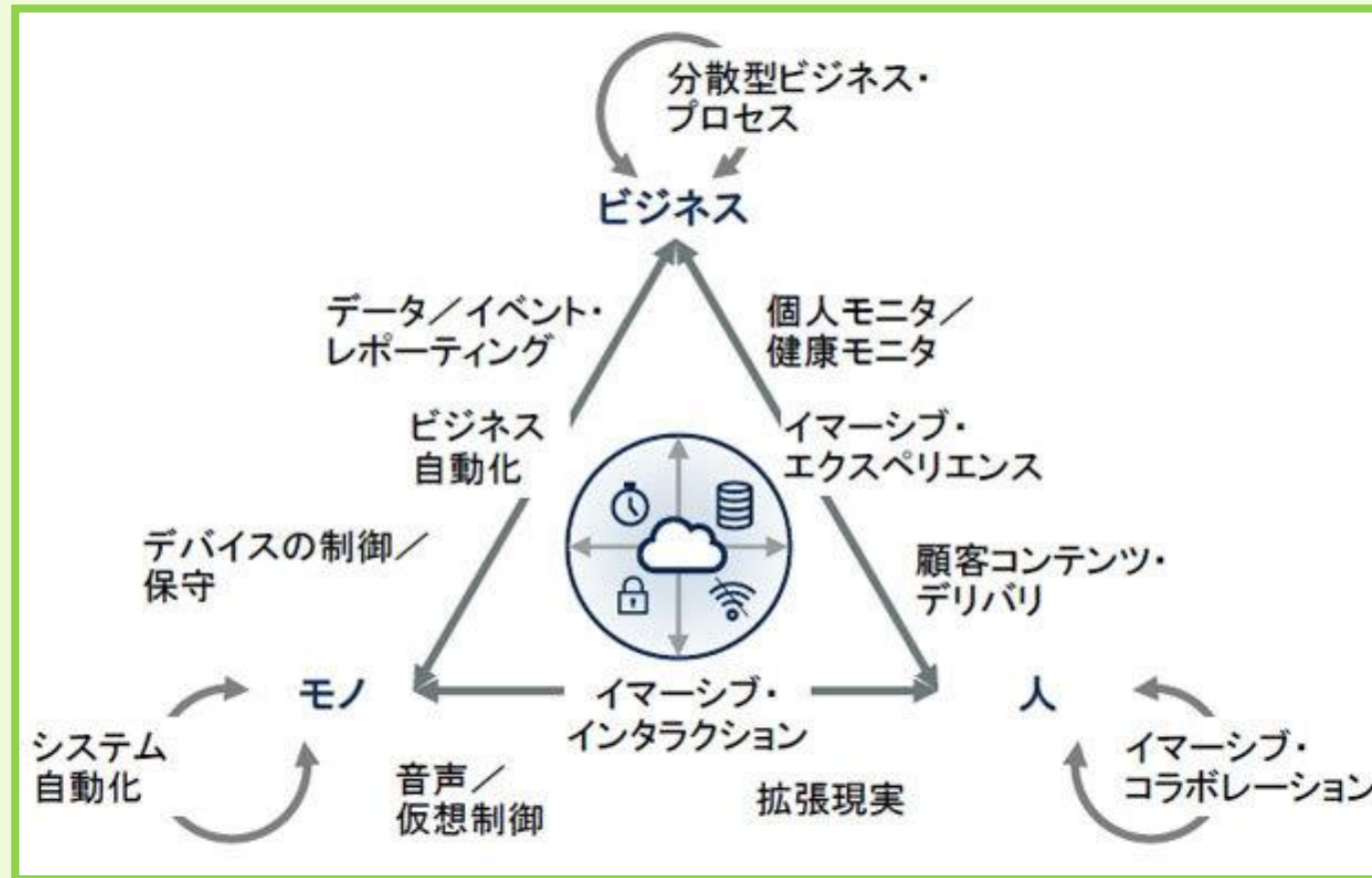
- ・ 工場の生産ラインで異常を検知し、その場で間引く

例3： 自動翻訳などの音声変換

- ・ 会話をリアルタイムで翻訳し音声化



エッジ・コンピューティングのユースケースは未知数であり、 ビジネスを起点にした12のカテゴリが参考になる



「エッジコンピューティングのユースケースにおける12のカテゴリ」
【出典】:Gartner



~Agenda~

1. エッジ・コンピューティングとは？

2. AIチップ

3. エッジAI製品紹介



「AIチップ」の性能がAIを活用した事業の競争力に直結

- 「AIチップ」とは？

- AIの計算処理に特化した半導体
- 「AIアクセラレータチップ」と呼ぶ場合も
- AI機能をハードウェア化するという発想



- 成長率1位の半導体は「車載向け」から「AIチップ」へ

- 2017～20年の3年間でAI向け半導体市場は2倍となる見込み！
- 2020～25年の5年間でさらに2倍となる予想も！



AIチップは学習用と推論用の2種類に大別される

- **[A]学習用**: クラウド側で使うサーバー機器向け

下記のような極めて高い演算能力が必要な処理が対象

- ディープラーニングによるAIの学習
- 学習済みAIによる推論を多数のユーザーに提供する場合

- **[B]推論用**: エッジ側の組み込み機器向け

- リアルタイムでの推論処理用
- より少ない消費電力で実行することが求められる



世界の主要企業がAIチップの開発に乗り出している①

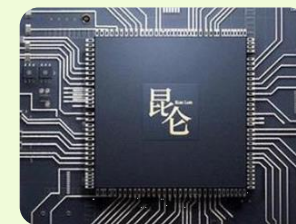
- Ø NVIDIA: **Tesla, DGX, Jetsonシリーズ**
- Ø Google: 「**TPU** (Tensor Processing Unit)」, v3(第3世代)
- Ø Intel: 「**NNP-L1000** (コードネーム: Spring Crest)」
- Ø Microsoft: 「Project Brainwave」 FPGAベース
- Ø Apple: 「**Neural Engine**」 A12 Bionic (69億個のトランジスタ)
- Ø Facebook: 自然言語処理に特化したAIチップ
- Ø Amazon: 「**AWS Inferentia**」 (買収したAnnapurna Labsの技術)
- Ø Tesla: 自律運転用新型チップ(「世界最高」と誇示)
- Ø IBM: WatsonからAIチップの開発へ!



世界の主要企業がAIチップの開発に乗り出している②

・AI大国に躍り出る中国 (**BATIS**):

- Ø **Baidu**: 「**Kunlun** (クンリン: 崑崙)」 [自動運転車分野]
- Ø **Alibaba**: 「**Ali NPU**」 [スマートシティ分野]
- Ø **Tencent**: [医療分野]
- Ø **iFlytek**(アイフライテック): 音声認識で世界トップレベル [音声認識分野]
- Ø **SenseTime**: [顔認識分野]



- Ø Huawei(ファーウェイ): 「**Kirin 980**」「**Ascend 910**」「**Ascend 310**」
- Ø DeePhi Tech: 「**DeePhi DPU**」 Xilinx社(ザイリンクス)に買収(2018/07)
- Ø Bitmain/Cambricon: ビットコインマイニング用AIチップ
- Ø Horizon Robotics



国内でもAIチップの研究開発は活発に行われている

・国内有名企業:

- Ø PFN (Preferred Networks): 「**MN-Core**」(2018年12月)
- Ø 富士通: 「**DLU** (Deep Learning Unit)」(「京」の技術を利用)
- Ø テンソーの半導体子会社 NSITEXE(エヌエスアイテクス):
自動車専用AIチップ「**DFP** (Data Flow Processor)」
- Ø ...



・スタートアップ企業:

- Ø AISing(岩手大学発のベンチャー): 「**AiiR(エアー)チップ**」(2019年1月)
- Ø Idein, ABEJA, ...



国内でもAIチップ開発事業を支援する動き

【NEDO】AIチップ開発加速のためのイノベーション推進事業

・事業期間: **2018～22年度(5年間)**

・2019年度予算: **16.8億円**

- ・ AIチップ等の開発には高度なスキルや高額な設計ツールが必要です。特に中小・ベンチャー企業にとっては、革新的なアイデアがあるにもかかわらず、新規参入等にあたりそれらが高いハードルになっています。そこで、中小・ベンチャー企業が持つアイデアを実用化するための設計開発を支援する事業を行います。
- ・ さらに、AIチップ等の開発を加速するために必要な共通基盤技術として、設計・評価・検証等の開発環境を有する拠点の整備、チップ開発を促進する共通技術の開発、IoTやAI技術を活用するための知見やノウハウを持った人材を育成する環境の整備を行います。

https://www.nedo.go.jp/activities/ZZJP_100142.html



エッジAIの高度化・実用化を推進するコンソーシアム「SCAiLE」

- 「**SCAiLE(スケイル)**」:

- 「SCalable AI for Learning at the Edge」(エッジ学習のためのスケーラブルなAI)

- 参画企業(日米4社): 2019年4月設立

- ① Crossbar社(米): 非常に高速な検索を実現する『Resistive RAM』技術

- ② Gyr Falcon Technology(米): AI処理を高速化する『AIアクセラレータ』技術

- ③ mtes Neural Networks(エムテス ニューラル ネットワークス: mtesNN社):
エッジデバイスやAIを結びつけるための技術とプラットフォーム

- ④ RoboSensing社(mtesNN社のグループ企業): ニューラルネットワーク関連技術



<http://prweb.jp/mtes/news-letter/3624>



AIチップ合戦、混戦から抜け出すのはどの企業か？

1. ユースケースのデファクト化:

- ・各業界をリードする使い手企業との共同開発
- ・産業応用でのフレイクスルー
- ・ユーザー目線（Apple:プライバシー保護への配慮を重視）

2. 注力先の見極め:

- ・学習向け or 推論向け
- ・プロセッサ(ASIC, FPGA, GPU, CPU)
- ・ポートフォリオ(カバレッジ)の広さ、〇〇専用チップ

3. 技術力の確保:

- ・低レイテンシ, コスト効率
- ・技術力の高い企業の買収

AI 川柳

画像処理 エヌビディアトップ 自動車
の トヨタと結び 自動運転

インテルは パソコン時代
王者なり 時代変転 今挑戦者

中国は 画像処理の 管理にて
8億市民 完全管理

IntelはオリンピックのAIプラットフォーム
パートナーで、2020年の東京オリンピック
ではAIを活用した新サービスを提供し
ていく計画らしいよ



~Agenda~

1. エッジ・コンピューティングとは？

2. AIチップ

3. エッジAI製品紹介



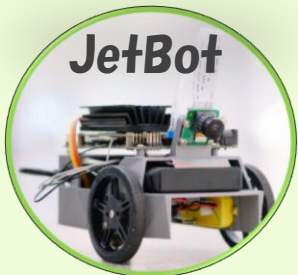
NVIDIA: Jetsonシリーズ





Jetson Nano(2019/04): \$99でGPUマシンが使える時代に！

Developer Kit



JetBot

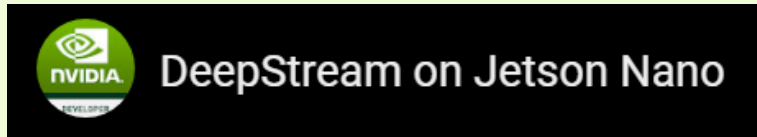


JETSON ファミリ エッジでのAIから自律動作マシンまで				
				
JETSON NANO 5 - 10W 0.5 TFLOPS (FP16)	JETSON TX1 → JETSON TX2 4 GB 7 - 15W 1 - 1.3 TFLOPS (FP16)	JETSON TX2 8GB Industrial 7 - 15W 1.3 TFLOPS (FP16)	JETSON AGX XAVIER 10 - 30W 10 TFLOPS (FP16) 32 TOPS (INT8)	
	エッジでのAI		完全自律マシン	
Nano (+/)	TX1	TX2	XAVIER (エクセビア)	
\$99	\$192	\$599	\$1,299	
2019年4月	2014年	2017年	2018年末	



NVIDIA: Jetson Nano

Jetson Nanoは最大8つのビデオストリームの並列リアルタイム処理が可能



<https://youtu.be/Y43W04sMK7I>



「Jetson Nano」はAI学習用教材にも採用！

- ・宇都宮大学がNVIDIAやFaBoと開発、授業への実践導入は世界初
- ・「AI ロボットカー」を利用した自動運転技術の習得を目指す



<https://robotstart.info/2019/05/29/utsunomiya-u-jetson.html>

NVIDIAコメント:

「Jetson Nanoが3月のGTC 2019で発表されましたが、学習教材としての採用は世界初になります。

日本の教育現場をこのような形で支援できることを大変喜ばしく思います」



【参考】2019年に注視すべき5つのAIトレンド (2019/01/01)

1. AIチップの勃興

- ・ 画像処理・自然言語処理・声認識など特定のユースケースに最適化されたAIチップも開発
- ・ 「FPGA」や「ASIC」をベースにしたAIチップの開発を強化する企業も！

2. エッジにおけるIoTとAIの融合

- ・ IoTデバイスから近い場所でコンピューティングを行う、エッジコンピューティングの重要性がさらに増す
- ・ 産業機器の異常検知・予知保全といった産業IoTにおいて利用されるケースがさらに増える

3. ONNX(オニクス)が相互運用性の鍵に

- ・ ONNXにより、TensorFlowやChainerなど、異なるDLフレームワーク間での移植が可能
- ・ AIエコシステムの主要プレーヤーは、推論用の標準ランタイムとしてONNXを使うようになる可能性

4. 機械学習の自動化が発展

- ・ データ分析者は、機械学習モデルの開発を自動化する「AutoML」を利用することでプロセスやワークフローで迷うことなく、ビジネス上の課題解決に専念することができる

5. 「AIOps」による自動化が進行

- ・ AIのITオペレーションやDevOpsへの導入は、組織にインテリジェンスをもたらし、オペレーションチームは精度の高い分析を行うことが可能になる

<https://forbesjapan.com/articles/detail/24693>



さいごに:

**エッジ・コンピューティングは
2019年の流行語
になるでしょうか？**



End of Document

