

The DENSO logo is written in a bold, italicized, red sans-serif font.

**DENSO**

Crafting the Core

# 機械学習モデルを搭載した セーフティクリティカルなシステムの 品質保証

桑島洋、安岡宏俊、中江俊博

株式会社デンソー

東京支社 電子基盤先行開発室

2018年08月31日

# 1.

背景

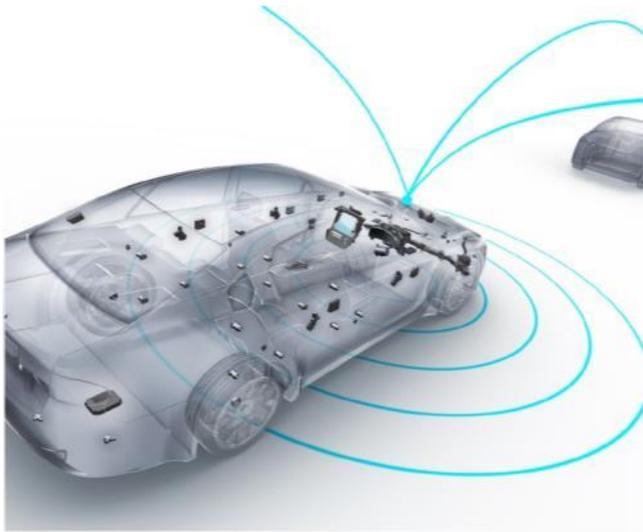


## 会社概要

設立		1949年12月16日
資本金		1,874 億円
売上収益	連結	5 兆 1,083 億円
営業利益	連結	4,127 億円
従業員数（就業人員ベース）	連結	168,813 名
	単独	39,315 名
連結子会社数		221 社（国内72・海外149）
持分法適用関連会社数		38 社（国内14・海外24）

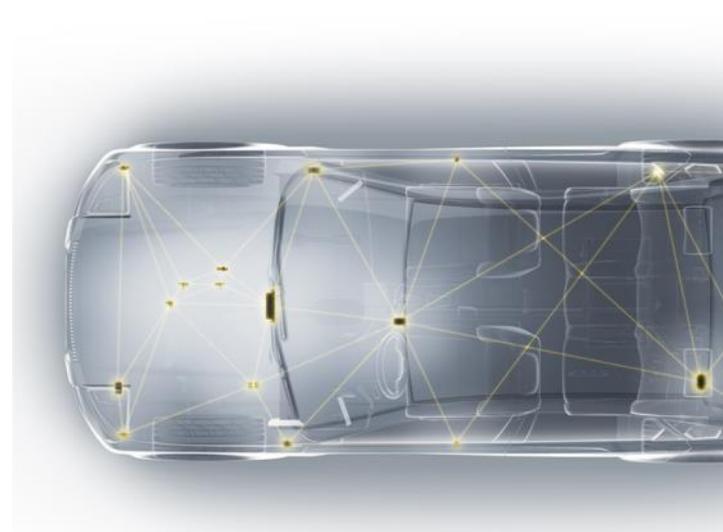
2018年3月31日現在

# 主な製品：自動車分野 1



## モビリティシステム

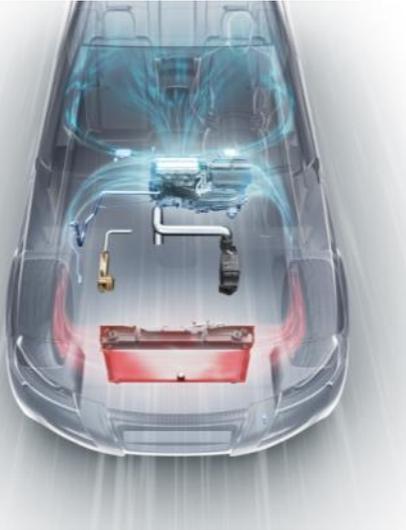
ミリ波レーダ、画像センサ、ドライバステータスマニタ、エアバッグシステム等の自動運転・先進安全製品、テレマティクスコントロールユニット、車車間・路車間通信機、ヘッドアップディスプレイ、メータ、コックピットシステム等のコネクティッド・コックピット製品



## 電子システム

車載向け I C・半導体センサ等のマイクロエレクトロニクスデバイス、エンジン制御コンピュータ・ボデー制御コンピュータ等のエレクトロニクス製品

# 主な製品：自動車分野 2



## サーマルシステム

自動車・バス用エアコンシステム、  
トラック用冷凍機、空調関係製品、  
ラジエータ等の冷却用製品



## パワートレインシステム

ガソリン・ディーゼルエンジン  
制御システム、燃料ポンプ、  
VVT、センサ等のエンジン関係  
製品、油圧制御バルブ等の駆動  
関係製品



## エレクトリフィケーション システム

HV・EVなどの電動化  
システム及び製品、  
オルタネータ、スタータなどの  
電源供給・始動システム製品、  
ワイパシステム、パワーウィンド・  
エアコン用等のモータ

# 主な製品:非車載事業

## 生活関連機器・産業機器



住宅設備機器



業務用空調機器



AUTO-ID関連機器



FA機器 (ロボット)

## 新事業分野



エネルギー  
マネジメント



農業 (AgTech)



農業 (AgTech)  
[コールドチェーン]



セキュリティ



情報ソリューション



ヘルスケア

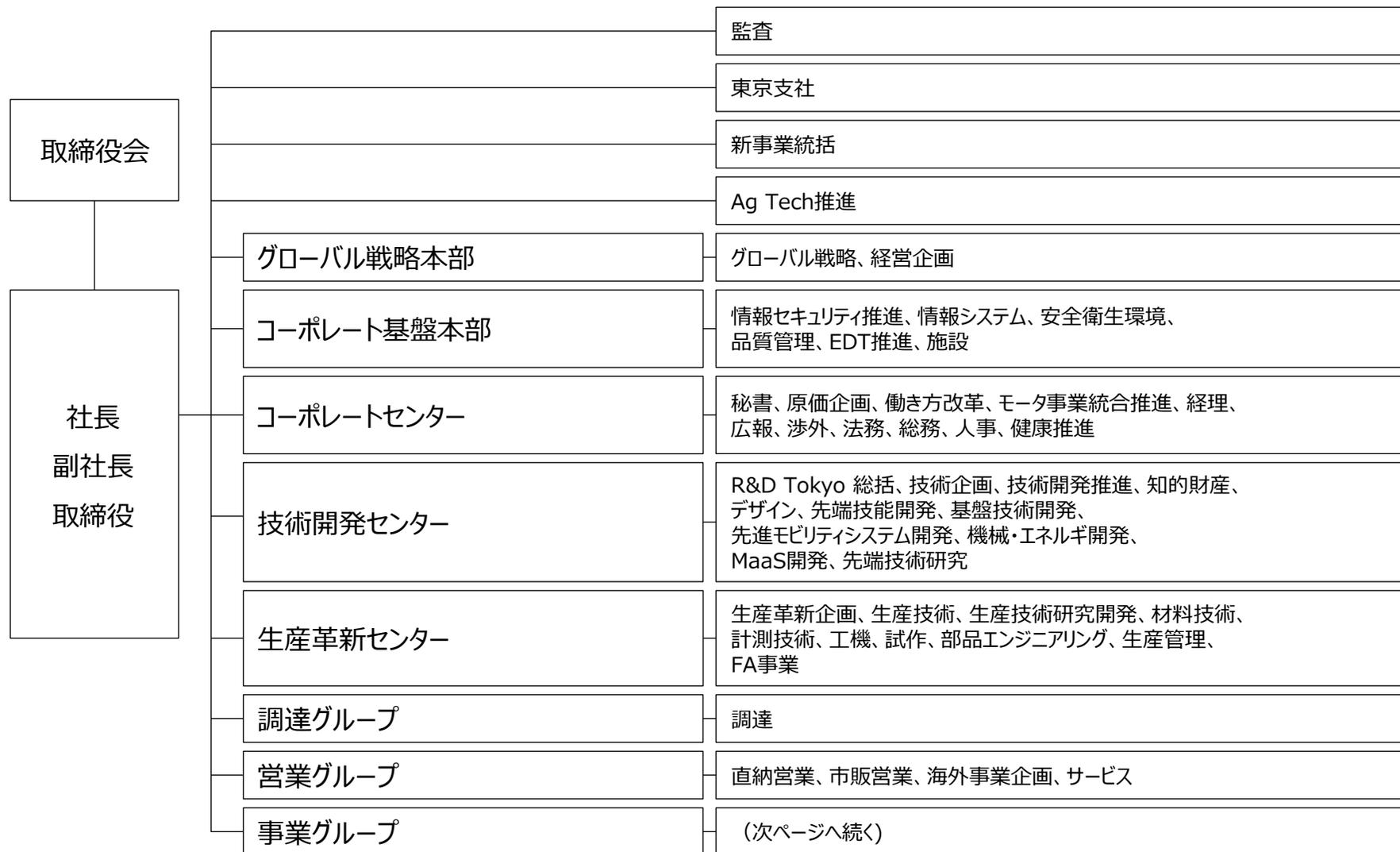


バイオ (微細藻類)



橋梁点検・測量サービス

# 会社組織



2018年6月20日現在

# 事業グループ（前ページの続き）

パワトレインシステム 事業グループ	エンジン制御システムおよび吸気系・燃料噴射系・点火系・後処理系・熱マネジメント系のシステム構成製品の開発・製造。 油圧制御バルブ・シフトバイワイヤ等の駆動系製品の開発・製造。
エレクトリフィケーションシステム 事業グループ	EV、ハイブリッド車等の電動化システムおよび製品の開発・製造。 オルタネータやスタータ等の電源供給・始動システム製品の開発・製造。 ワイパシステムやスマートモータ等の製品の開発・製造。
電子システム事業グループ	半導体センサ、IC等のマイクロエレクトロニクスデバイス製品の開発・製造。 エンジン制御コンピュータ、ボデー制御コンピュータ等のエレクトロニクス製品の開発・製造および電子基盤技術開発。
サーマルシステム事業グループ	乗用車用エアコンシステム、快適性システムと製品の開発・製造。 パワトレイン冷却システムと製品の開発・製造。
モビリティシステム事業グループ	モビリティ全体電子システム、サービス、プラットフォームの開発・提供。 ミリ波レーダ、画像センサ、ドライバステータスマニタ、エアバッグシステム等の <b>自動運転・先進安全製品の開発・製造</b> 。 テレマティクスコントロールユニット、車車間・路車間通信機、ヘッドアップディスプレイ、メータ、コックピットシステム等の <b>コネクティッド・コックピット製品の開発・製造</b> 。

2018年6月20日現在

# AIとは

## AI:人間の知的行動をコンピュータ化した技術

### AIの活用分野

#### 不動産査定

推論予測に活用



【現地】高架の橋梁など



画像送信

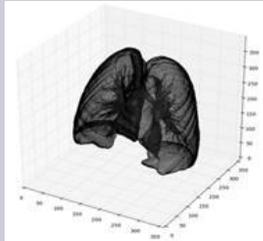
診断結果

【クラウド】ひび割れ検出、損傷判別 画像診断システム



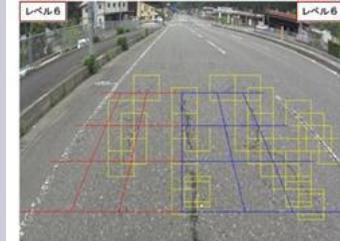
#### 医療・創薬

推論予測に活用



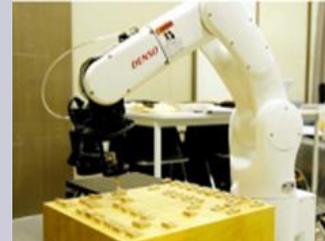
#### インフラ点検

画像認識に活用

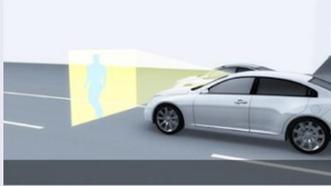
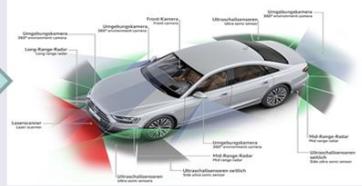
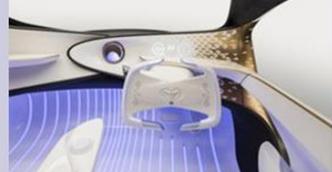


#### 囲碁将棋

行動計画に活用



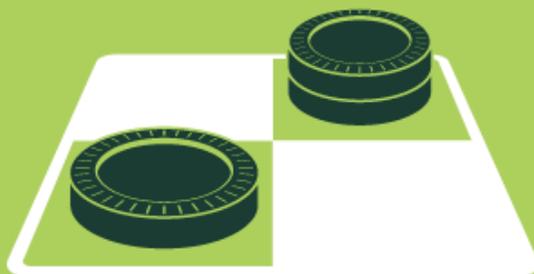
# 自動車へのAI活用例

活用領域	活用領域	
	現在	今後
安心	<p><b>予防安全製品の画像認識</b></p>  <p><a href="http://www.denso-recruit.jp/company/safety.html">http://www.denso-recruit.jp/company/safety.html</a></p> <p>例)プリクラッシュセーフティ</p>	<p><b>自動運転における判断/操作</b></p>  <p><a href="http://jakopnik.com/the-new-semi-autonomous-audi-a8-will-let-drivers-watch-179680629">http://jakopnik.com/the-new-semi-autonomous-audi-a8-will-let-drivers-watch-179680629</a></p> <p>例)Audi A8 自動運転Lv3</p>
	<p><b>音声認識による目的地設定</b></p>  <p><a href="http://www.denso.co.jp/ai/news/newsreleases/2013/130925-02.html">http://www.denso.co.jp/ai/news/newsreleases/2013/130925-02.html</a></p> <p>例)ナビ</p>	<p><b>感情推定・ニーズ予測</b></p>  <p><a href="http://www.toyota.co.jp/jpn/tech/smart_mobility_society/concept-ii">http://www.toyota.co.jp/jpn/tech/smart_mobility_society/concept-ii</a></p> <p>例)トヨタ CONCEPT-愛i</p>
利便	<p><b>規定区間内輸送</b></p>  <p><a href="http://www.expo2005.or.jp/jp/AI/AS/AS-1/">http://www.expo2005.or.jp/jp/AI/AS/AS-1/</a></p> <p>例)IMTS IMTS: Intelligent Multimode Transit System</p>	<p><b>任意区間輸送</b></p>  <p><a href="http://japan.cnet.com/article/05087850/">http://japan.cnet.com/article/05087850/</a></p> <p>例)ロボットタクシー</p>

# AI、機械学習、ディープラーニング

## 人工知能 (AI)

初期の AI が注目を集める



## マシンラーニング (機械学習)

機械学習が活発化し始める



## ディープラーニング (深層学習)

ディープラーニングのブレイクスルーが AI ブームを巻き起こす



1950's

1960's

1970's

1980's

1990's

2000's

2010's

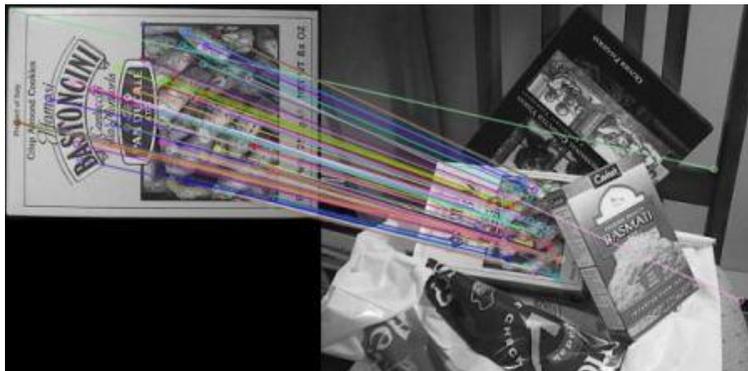
1950年代の楽観主義が失望に変わって以来、最初は人工知能のサブセットである機械学習、続いて機械学習のサブセットであるディープラーニングが、これまでにない破壊的イノベーションを起こしています。

<https://blogs.nvidia.co.jp/2016/08/09/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/>

# ディープラーニングによるブレイクスルー

## 画像認識の場合

ディープラーニング  
以前



[https://docs.opencv.org/3.3.0/dc/dc3/tutorial\\_py\\_matcher.html](https://docs.opencv.org/3.3.0/dc/dc3/tutorial_py_matcher.html)

顔検出：識別器のラストスキャン



Haar-like 特徴

AdaBoost

弱識別器 $h$	重み $\alpha$
$h_1(\cdot)$	$\times \alpha_1 \rightarrow 0.5$
$h_2(\cdot)$	$\times \alpha_2 \rightarrow 0.3$
$\vdots$	$\vdots$
$h_r(\cdot)$	$\times \alpha_r \rightarrow 0.1$
	$\parallel$
	<b>2.3</b>

顔 / 非顔を判定

[https://www.slideshare.net/MPRG\\_Chubu\\_University/ss-32258845](https://www.slideshare.net/MPRG_Chubu_University/ss-32258845)

27

2012年頃

ブレイクスルー

局所的なパターンマッチング

汎化能力の壁

特定のドメイン

ディープラーニング  
以後



[https://d1fshsy4d47upe.cloudfront.net/35/35ce0cedf846f526268298bed40759bb\\_t.jpeg](https://d1fshsy4d47upe.cloudfront.net/35/35ce0cedf846f526268298bed40759bb_t.jpeg)



[https://d1fshsy4d47upe.cloudfront.net/bd/bd1d318d8f8a0bd6ee4975f5aa96a2d3\\_t.jpeg](https://d1fshsy4d47upe.cloudfront.net/bd/bd1d318d8f8a0bd6ee4975f5aa96a2d3_t.jpeg)



<http://sozaing.com/wp-content/uploads/pizza-480x360.jpg.pagespeed.ce.AMbUoHvdvZ.jpg>

“ピザ”



<https://www.pakutaso.com/20160953245post-8887.html>



<https://publicdomainq.net/images/201710/18s/publicdomainq-0014503pxb.jpg>

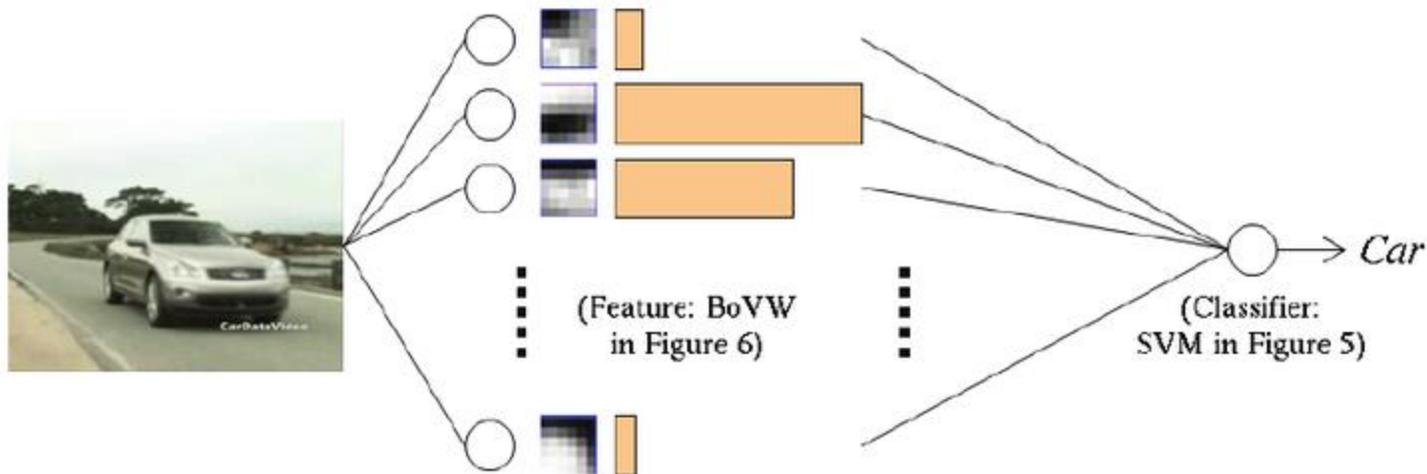


<https://www.pakutaso.com/20160953245post-8887.html>

“お好み焼き”

# ディープラーニング（深い学習）の汎化能力の秘密

a) Traditional machine learning approach using a hand-crafted feature



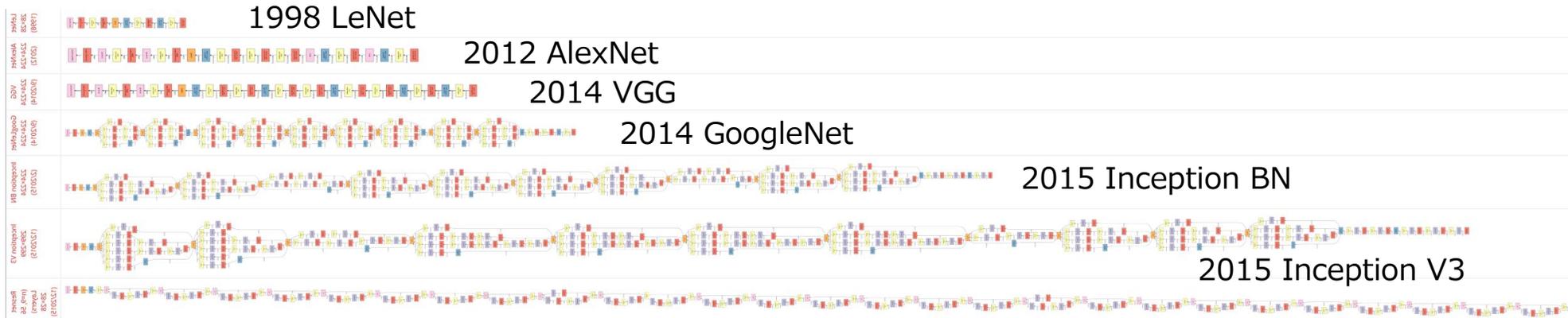
b) Deep learning approach



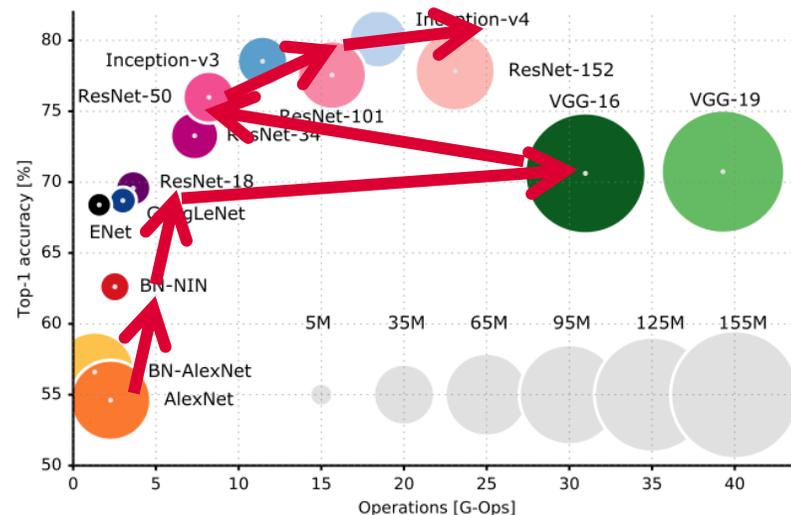
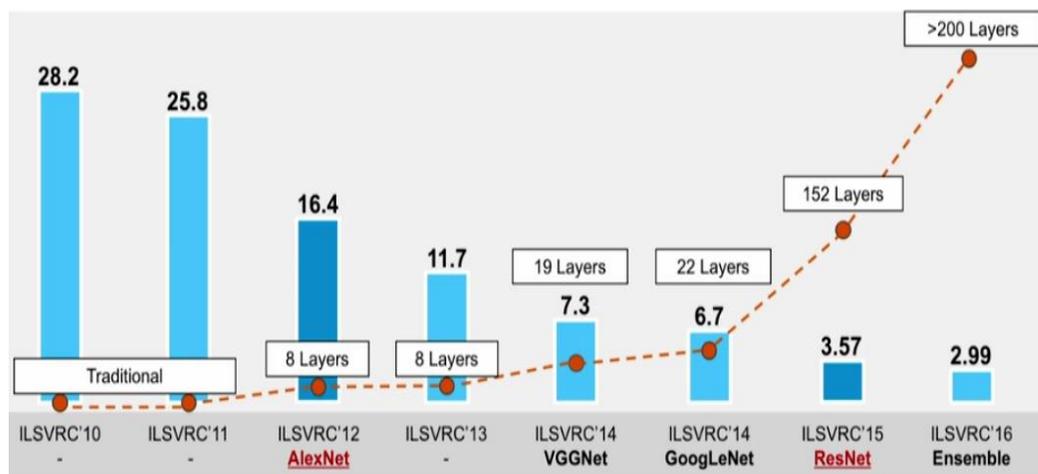
[https://www.researchgate.net/figure/A-conceptual-comparison-between-traditional-machine-learning-and-deep-learning\\_fig8\\_282203673](https://www.researchgate.net/figure/A-conceptual-comparison-between-traditional-machine-learning-and-deep-learning_fig8_282203673)

[http://www.slideshare.net/NVIDIA/visual-computing-the-road-ahead-an-nvidia-ces-2015-presentation-deck/30-30HOW\\_A\\_DEEP\\_NEURAL\\_NETWORK](http://www.slideshare.net/NVIDIA/visual-computing-the-road-ahead-an-nvidia-ces-2015-presentation-deck/30-30HOW_A_DEEP_NEURAL_NETWORK)

# 層数の増加 → 認識精度向上 → ブラックボックス化加速



<http://josephcohen.com/w/visualizing-cnn-architectures-side-by-side-with-mxnet/>



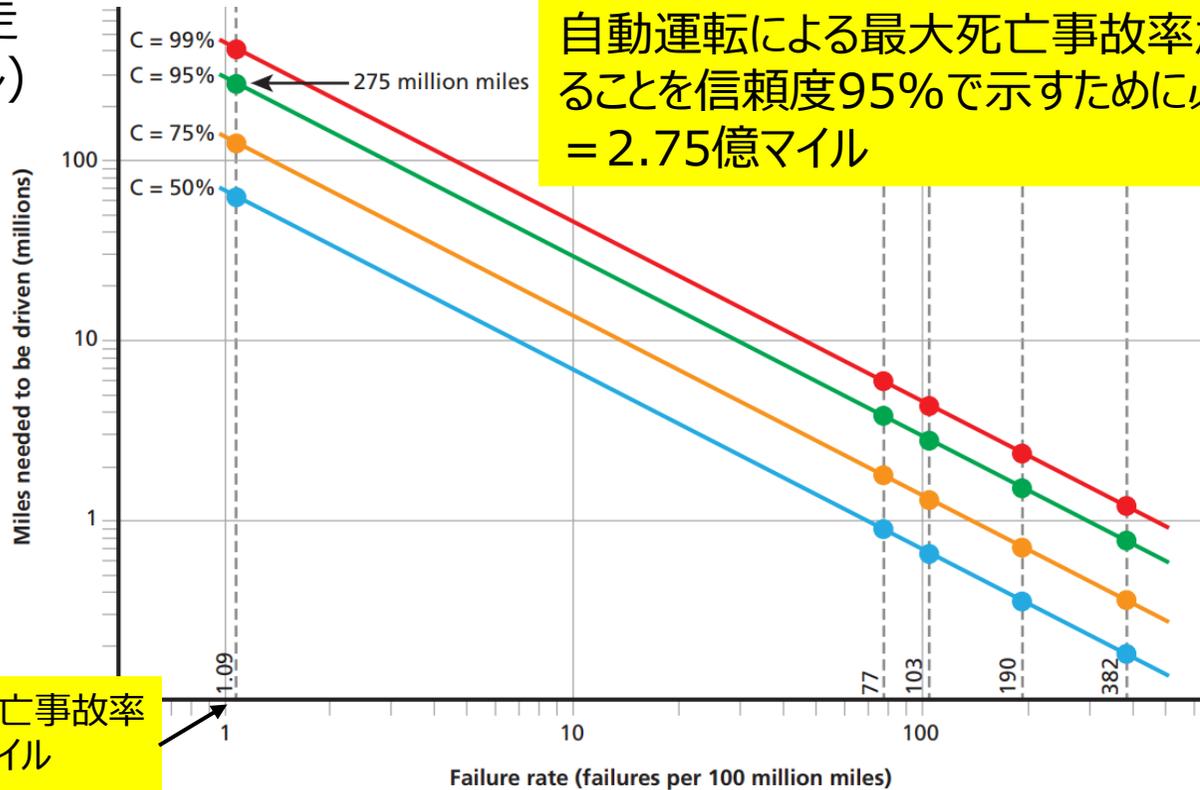
<https://www.quora.com/Is-wider-or-deeper-convolutional-neural-networks-better-for-computer-vision-tasks>

[https://www.researchgate.net/figure/Comparison-of-popular-CNN-architectures-The-vertical-axis-shows-top-1-accuracy-on\\_fig2\\_320084139](https://www.researchgate.net/figure/Comparison-of-popular-CNN-architectures-The-vertical-axis-shows-top-1-accuracy-on_fig2_320084139)

# 膨大なデータが必要 → 現実的な検証の指針なし

Figure 1. Failure-Free Miles Needed to Demonstrate Maximum Failure Rates

縦軸：必要走行距離(マイル)



自動運転による最大死亡事故率が1.09件/1億マイルであることを信頼度95%で示すために必要な無事故走行距離 = 2.75億マイル

人の運転による死亡事故率 = 1.09件/1億マイル

横軸：1億マイル当たりの事故件数

SOURCE: Authors' analysis.

NOTE: The four colored lines show results for different levels of confidence. The five dashed vertical reference lines indicate the failure rates of human drivers in terms of fatalities (1.09), reported injuries (77), estimated total injuries (103), reported crashes (190), and estimated total crashes (382).

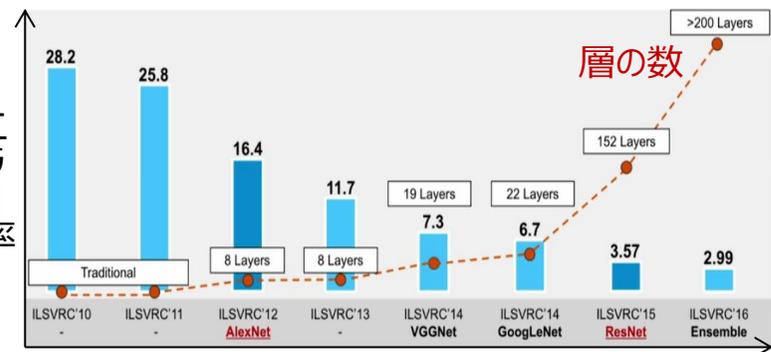
RAND RR1478-1

<https://www.autosafety.org/wp-content/uploads/2016/04/RAND-AV-Report.pdf>

# AI品質基盤に取り組む背景

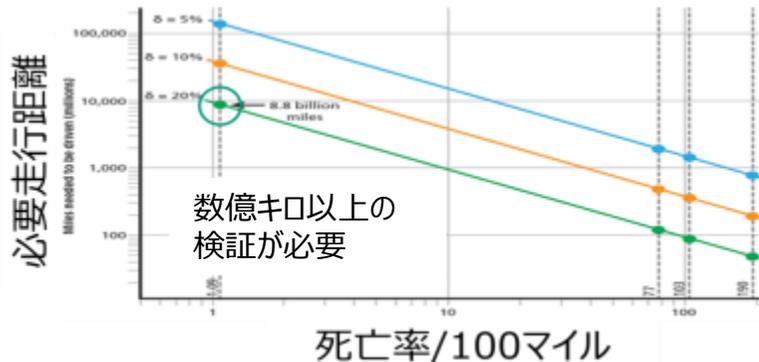
## 【背景】

推論精度が大幅向上⇒ブラックボックス化が加速



Deep Learningの進化

膨大なデータが必要⇒現実的な検証の指針なし



## 【進め方】

### 先進技術・事例

- ・学会、研究機関
- ・航空・宇宙など先進異分野 (主にDIAM SVIC連携)



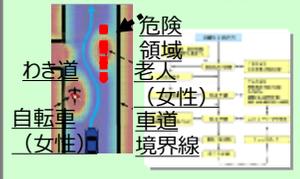
### 産官学動向

- ・欧州を中心としたプロジェクト
- ・国際標準化
- ・国内業界・団体 (主にDNDE CRD連携)



### 社内事例・計画

- ・開発プロセス
- ・設計評価技術 (社内関連部署と連携)



調査・俯瞰

課題打ち上げ

調査・分析

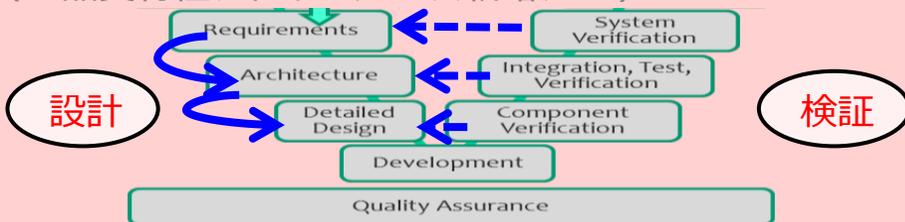
合意形成

調査・分析

適用・展開

## AI品質基盤：開発プロセスと基盤技術

(AI品質特性、オープンソース戦略、...)



課題認識はあるが対応手法は未確立

各種動向をもとに基盤開発、業界合意&社内適用を狙う

# AI品質基盤の考え方

**車両** ⇨ 業界・標準化動向を把握し、相場観を見極める

安全要件、検証方法の業界コンセンサスを探り開発指針とする

事例

**欧州: PEGASUS**

車両レベルのシナリオ検証が主体。高速道Lv3のシーンを9分類



**米国: NHTSA**

自動運転安全ガイドライン発行。12項目の自主報告を推奨



**他社: Mobileye**

事故責任概念を定式化した独自の安全ポリシー(RSS)とHDマップ(REM)を提唱

**システム** ⇨ AIをシステムレベルで保証する技術・技法を開発する e.g. AIコンポのブラックボックステスト手法

従来のソフトウェア工学が通用しない

ソフトウェア工学の考え方・手法をAI搭載システム向けに拡張する動き

新たな考え方 **技術が必要**

**システム品質**

仕様書                      プログラ                      ソースコード

```

// Simple HelloButton() method.
+ version 1.0
+ developer john doe <mailto:john.doe@company.com>
+ last modified 2023-01-01
+ use the JFrame type until support for t
+ don't use "Hello Button"
+ use the "Frame.getContentPane()"
// display the fra
    
```

国内	MLSE	機械学習工学研究会
	QA4AI	AI品質保証コンソーシアム
海外	SEMLA	海外版MLSE モントリオール工科大中心
	DISE	セーフティクリティカルシステムのためのDL@ICML
	WAISE	AIの安全工学ワークショップ @SafeComp
標準化	SOTIF	機能安全の拡張版

**機械学習モデル** ⇨ データ品質・モデル品質の両面から、AIの品質課題の根本的解決を目指す

**データ品質**

どれだけタグ付けすればよいのか分からない

大量データへのタグ付け作業が大変

タグの仕様の決め方が分からない

⋮

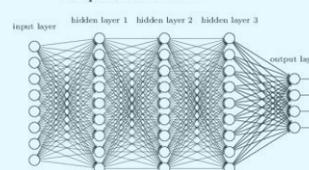


**データ**

大量データに人間が正解タグを付与したデータ

機械学習

Deep neural network



**学習済モデル**

数億ものパラメータで構成される判定ロジックを帰納的に構築

**モデル品質**

パラメータが膨大で判定結果の説明ができない

特定のノイズで簡単に間違えてしまう

テストの十分性が保証できない

⋮

**俯瞰**に裏打ちされたAI品質の考え方構築・技術開発・仕組み作り・業界打込み

# 2.

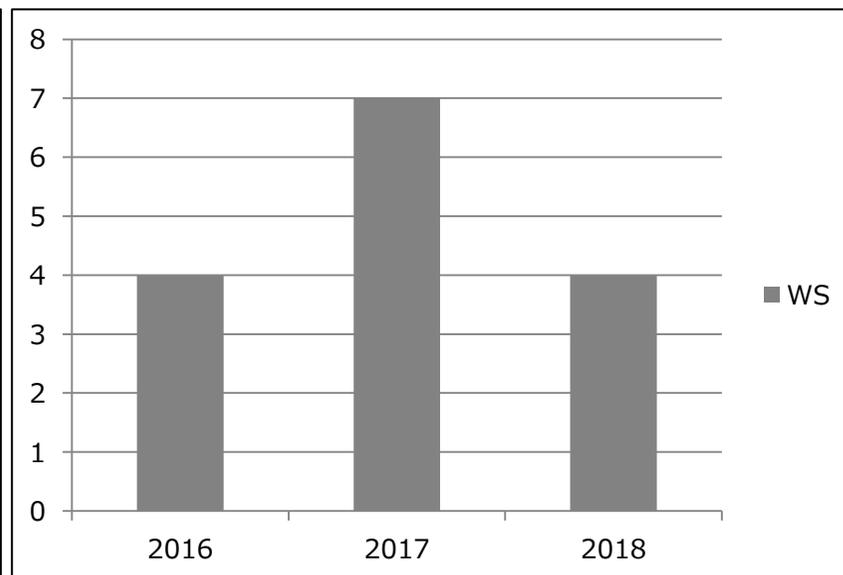
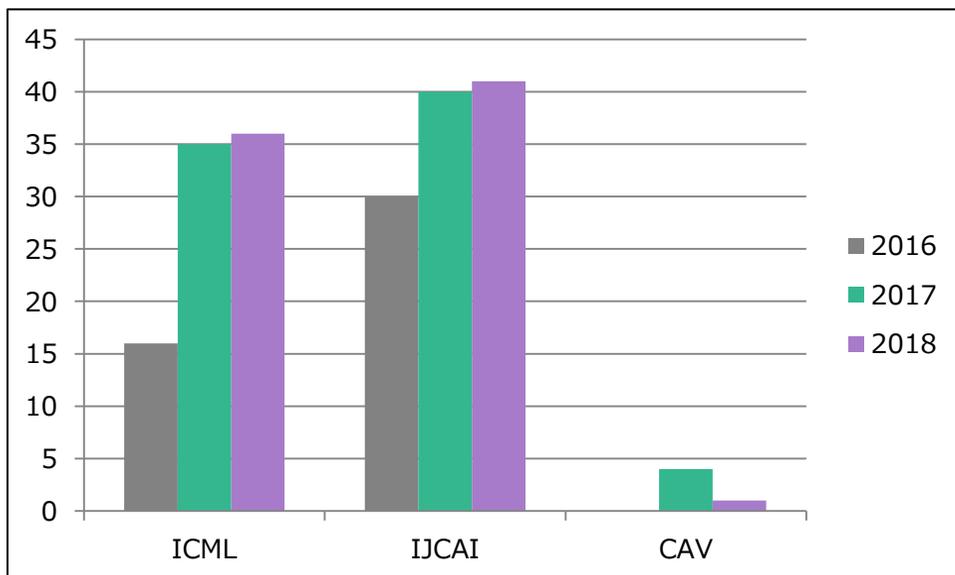
## 動向

# 学会の発表動向

AI品質に関連する発表数：

- 国際機械学習会議:ICML,
- 国際人工知能会議:IJCAI,
- コンピュータ支援検証に関する国際会議:CAV

関連WS数 (ICML+IJCAI併設)：



発表総数：ICML→約600件, IJCAI→約700, CAV→約60

アカデミアでのAI品質への関心が見える

# 論文のタイトルから見るトレンドの変化 (ICML2018)

【2017年】

【2018年】



## ICML2018ベストペーパー

Anish Athalye, Nicholas Carlini, David Wagner, "Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples" → **新たなAIの脆弱性を発見した研究**

新たなモデルの開発競争は少し落ち着き、性質の解明や応用にシフトしつつある

# Adversarial exampleとは

**Adversarial example** [Szegedy et al., ICLR'14] :

ニューラルネットワークモデルが誤判定するように、微小なノイズを足した入力。



?

ニューラルネットワークモデルと画像から、AEを作成する攻撃方法が研究されている

# 自動車に関連する Adversarial example (1/2)

## The Nexar traffic light challengeを基にしたAE

[Wicker et al., TACAS'18]:



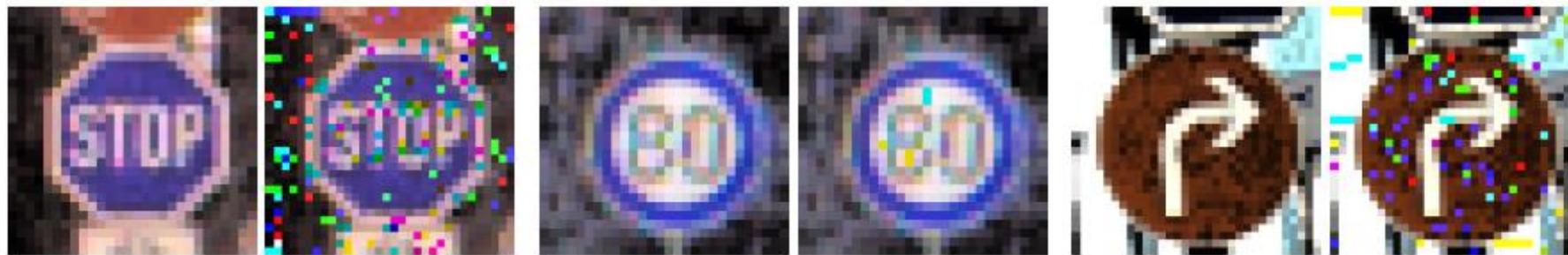
[Wicker et al., TACAS'18]

1ピクセル画像を変更することで、信号の判定が失敗する

# 自動車に関連する Adversarial example (2/2)

## German Traffic Sign Recognition Benchmarkを基にしたAE

[Huang et al., CAV'17]:



ストップ  
⇒ 30m速度制限

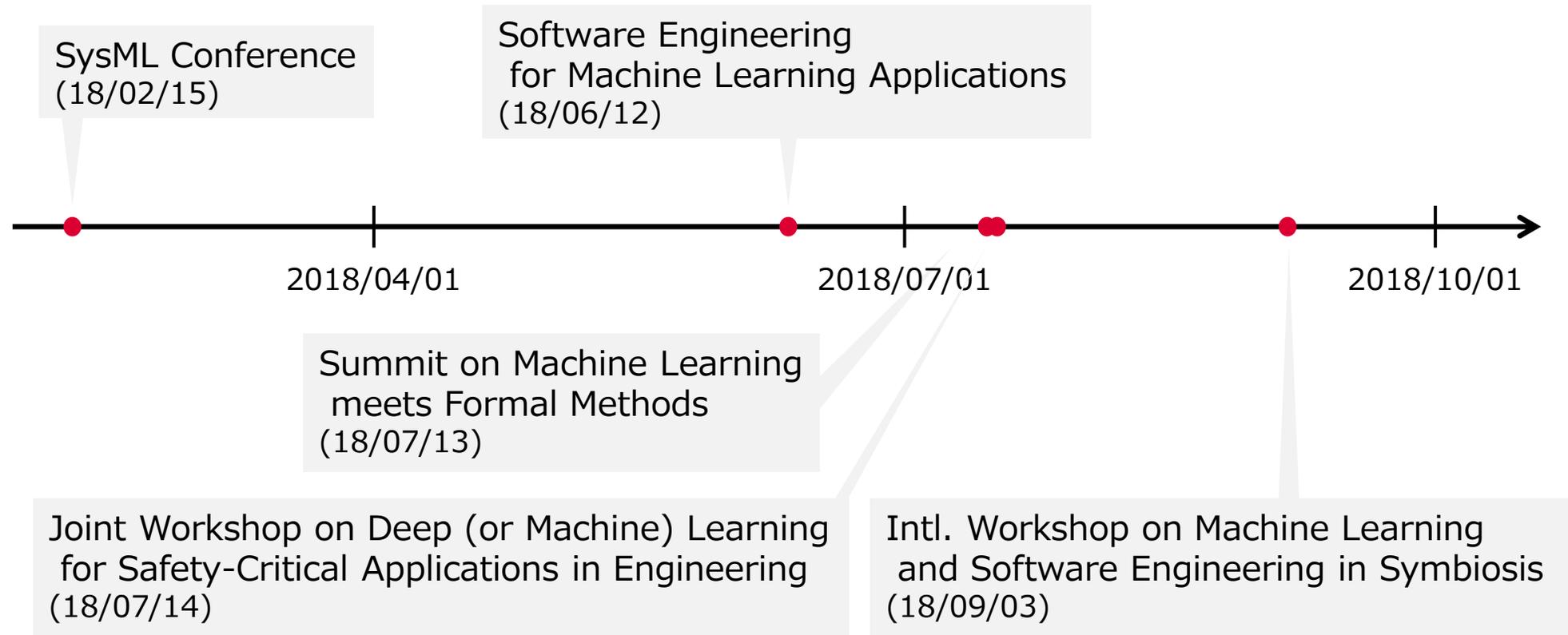
80m速度制限  
⇒ 30m速度制限

右折指示  
⇒ 直進指示

[Huang et al., CAV'17]

数ピクセル画像を変更することで、交通標識判定が失敗する

# 機械学習に対するソフト/システム工学 関連会議



アカデミアでのAI品質への関心が見える

# 機械学習工学研究会

- 2018年4月 日本ソフトウェア科学会から正式な研究会として発足
- 機械学習システムの開発・テスト・運用の方法論の体系化が設立趣旨
- これまでの主な活動と今後の予定
  - 5月 キックオフシンポジウム
  - 7月 第1回機械学習工学研究会ワークショップ (MLSE2018)
  - 8月 日本ソフトウェア科学会大会
  - 8月 機械学習工学研究会 現場を交えた勉強会 #1



Gold Sponsors



Silver Sponsors



Bronze / Special Sponsor



キックオフシンポジウムのスポンサー

# MLSE2018（機械学習工学研究会ワークショップ）

## 1日目：ポジション発表

- 桑島, 安岡, 中江(デンソー), “セーフティクリティカルな機械学習システムの品質保証の課題”
- 石川冬樹(NII), 松野裕(日大), “Continuous Argument Engineering：機械学習システムにおける議論モデルの活用に向けて” →議論モデルで考慮すべき機械学習固有の不確かさを定義
- 明神(JAXA), “機械学習応用システムの信頼性に関する一考察” →従来の信頼性の定義をもとに機械学習応用システムで考慮すべき信頼性を考察

など、22件の発表

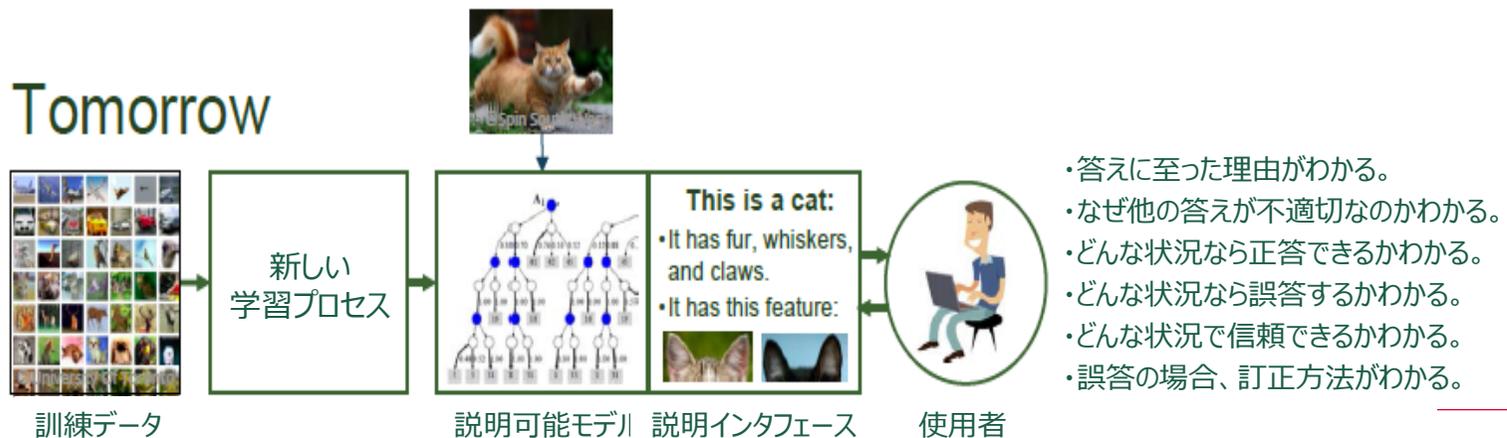
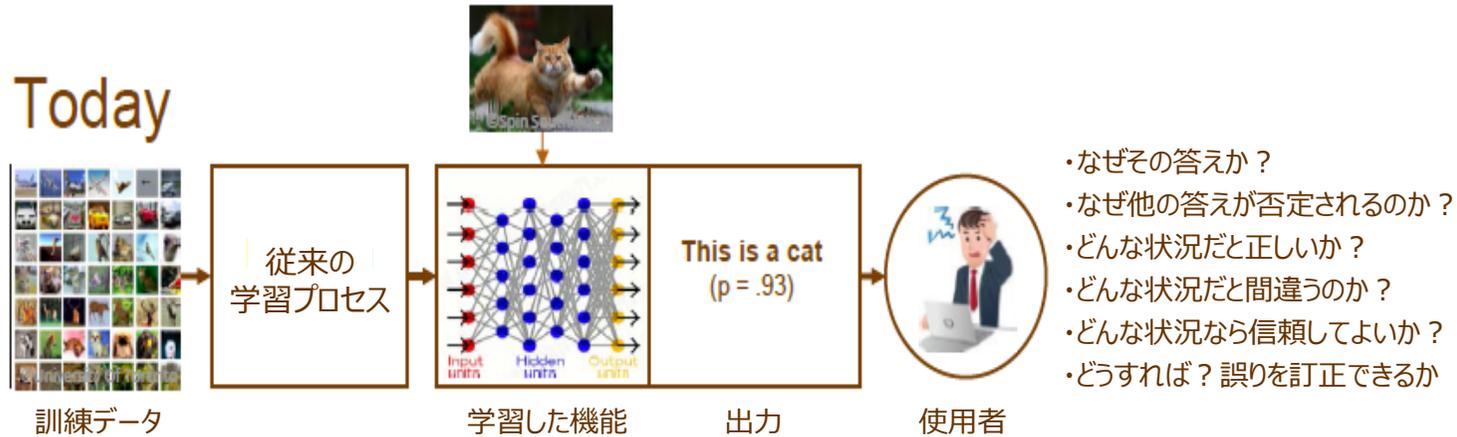
## 2日目：議論されたテーマ

大分類	テーマ名
テスト手法（4テーマ）	・DNN+画像テスト① ・DNN+画像テスト② ・テスト全般 ・ML開発実務者の自動テスト設計ベストプラクティス
評価指標（4テーマ）	・機械学習システムの評価 ・AIと品質特性 R・A・M・S・Sec ・不確かさとは？ ・モデル圧縮
データ/開発環境（5テーマ）	・データ管理 ・こんなデータはゴミだ ・技術的負債 ・機械学習の開発・実行基盤・再利用
工数見積もり/運用（4テーマ）	・工数を見積もる ・MLシステムの運用とヒト ・運用容易性 ・サービスレベルと運用保守体制
調査/体系化（8テーマ）	・BOKの検討 ・プロセス ・国からのファンディング ・人材育成組織論 ・調査 ・コミュニティの作り方 ・「酒屋問題」的な標準例題 ・MLSEのこれからを話そう

機械学習の開発・運用環境、機械学習の評価、体系化/BOK化 が主な関心事

# 国防高等研究計画局 (DARPA) 説明可能AI (XAI)

AIシステムの**意思決定や行動をユーザに説明できない**という問題点を克服し、人工知能をパートナーとして理解し、正しく信用し、効率的に管理するため、説明可能 (Explainable) な人工知能 = **XAI**を開発する。



# DARPA XAI : プログラムマネージャ・パフォーマーと日程



テキサスA&M大



ルトガス大



UCバークレー

charles river analytics

チャールズリバー



UCロサンゼルス



UTダラス



Mr. David Gunning  
Information Innovation Office (I20)  
Program Manager

DARPA



Oregon State  
University

オレゴン州立大



レイセオンBBN



スタンフォード研究所



カーネギーメロン大



A Xerox Company

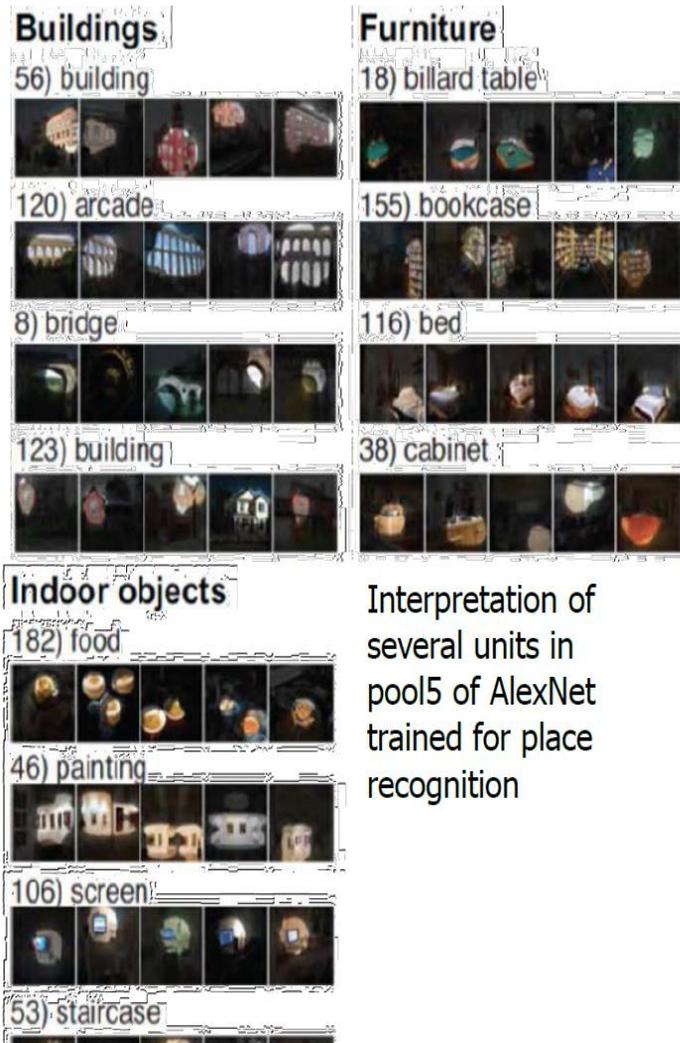
ゼロックス  
パロアルト研究所

## プロジェクト期間

キックオフ : 2017年5月 / Phase 1 : ~2018年10月 / Phase 2 : ~2021年4月

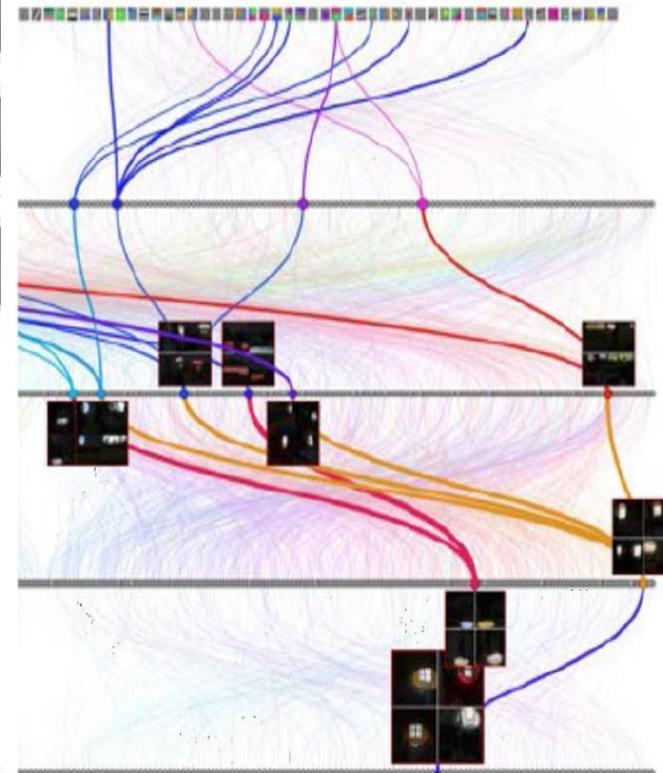
XAIには主要大学と企業研究所が多数参加

# DARPA XAI : Network Dissection (MIT)



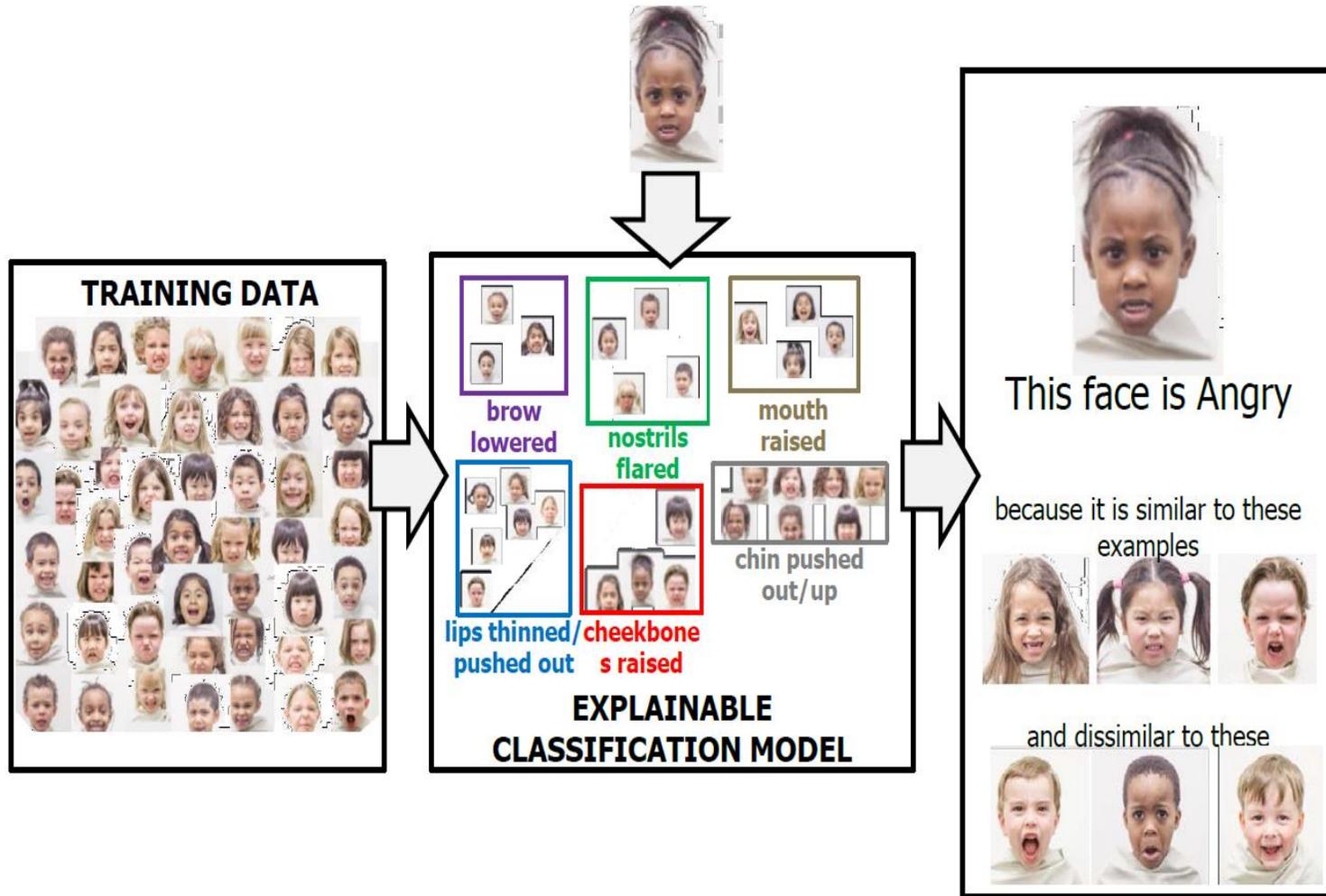
Interpretation of several units in pool5 of AlexNet trained for place recognition

Audit trail: for a particular output unit, the drawing shows the most strongly activated path



ニューラルネットワークの各ノードが検出しているオブジェクトが見える化

# DARPA XAI : Selection of Teaching Ex (RUT)



判定結果とともに、根拠（類似事例と非類似事例）を提示

# 3.

## デンソーのAI品質基盤が考える課題

# AI品質基盤の考え方（再掲）

**車両** ⇨ 業界・標準化動向を把握し、相場観を見極める

安全要件、検証方法の業界コンセンサスを探り開発指針とする

事例

**欧州: PEGASUS**

車両レベルのシナリオ検証が主体。高速道Lv3のシーンを9分類



**米国: NHTSA**

自動運転安全ガイドライン発行。12項目の自主報告を推奨



**他社: Mobileye**

事故責任概念を定式化した独自の安全ポリシー(RSS)とHDマップ(REM)を提唱

**システム** ⇨ AIをシステムレベルで保証する技術・技法を開発する e.g. AIコンポのブラックボックステスト手法

従来のソフトウェア工学が通用しない

ソフトウェア工学の考え方・手法をAI搭載システム向けに拡張する動き

新たな考え方 技術が必要

システム品質

仕様書      プログラ      ソースコード

```

// Simple HelloButton() method.
+ version 1.0
+ developer john doe <mailto:john.doe@company.com>
+ last modified 2017-01-01
+ use the JFrame type until support for t
+ content is finished
+ title is "Hello Button"
+ frame.getContentPane().
// display the fra
    
```

国内	MLSE	機械学習工学研究会
	QA4AI	AI品質保証コンソーシアム
海外	SEMLA	海外版MLSE モントリオール工科大中心
	DISE	セーフティクリティカルシステムのためのDL@ICML
	WAISE	AIの安全工学ワークショップ @SafeComp
標準化	SOTIF	機能安全の拡張版

**機械学習モデル** ⇨ データ品質・モデル品質の両面から、AIの品質課題の根本的解決を目指す

## データ品質

どれだけタグ付けすればよいのか分からない

大量データへのタグ付け作業が大変

タグの仕様の決め方が分からない

⋮

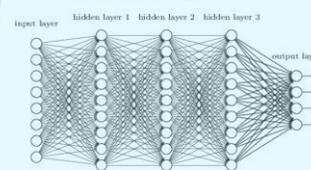


## データ

大量データに人間が正解タグを付与したデータ

## 機械学習

Deep neural network



## 学習済モデル

数億ものパラメータで構成される判定ロジックを帰納的に構築

## モデル品質

パラメータが膨大で判定結果の説明ができない

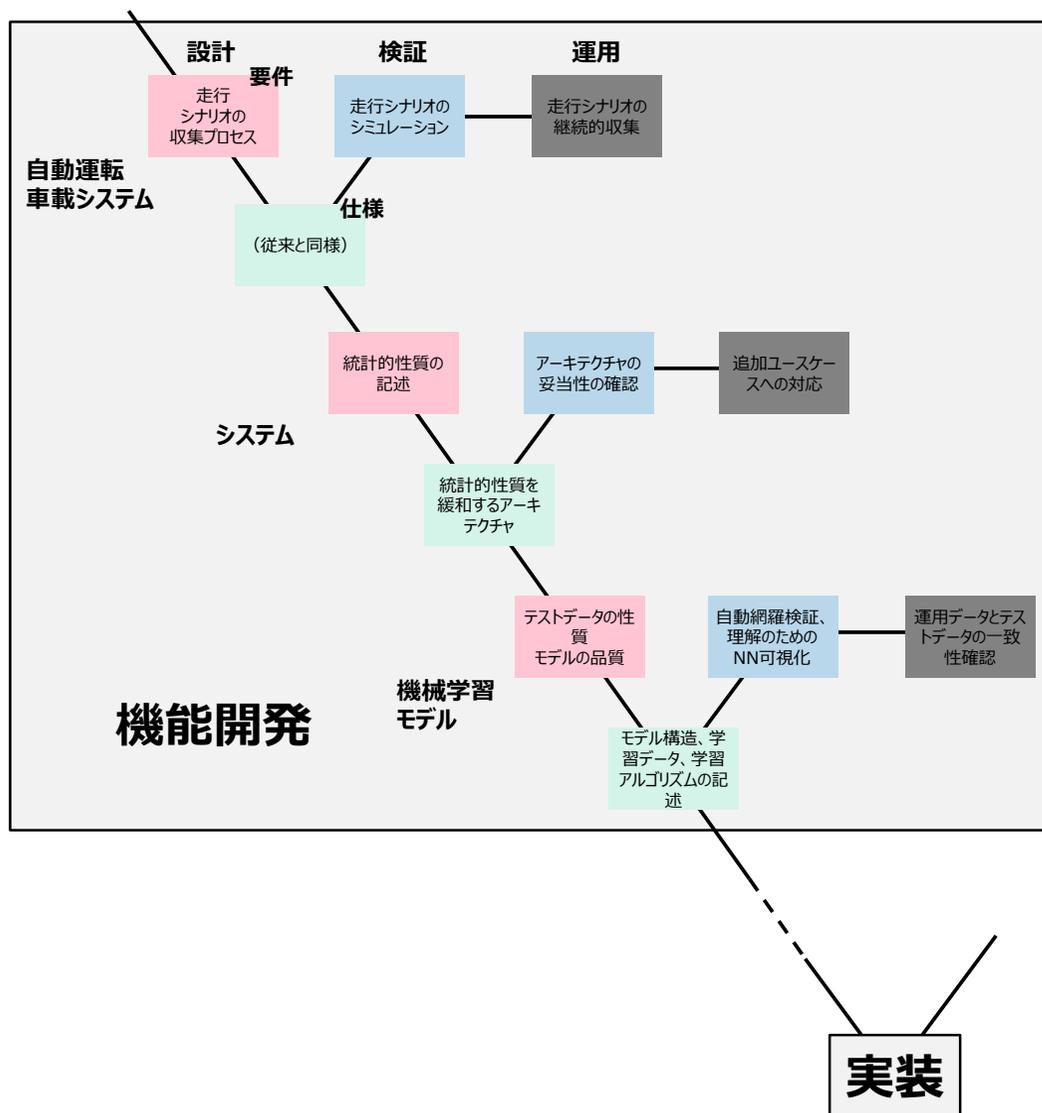
特定のノイズで簡単に間違えてしまう

テストの十分性が保証できない

⋮

俯瞰に裏打ちされたAI品質の考え方構築・技術開発・仕組み作り・業界打込み

# 課題の位置づけ



本日紹介するのは、機械学習システムの機能開発における課題（実装の課題はまだ明らかでない）

# 車両における機械学習システム品質保証の課題

車両レベル

→ **物理環境に接する階層**

理論的（解析的）に保証することは難しく、  
実験的に獲得した**物理環境の統計的性質**  
を反映することが重要。

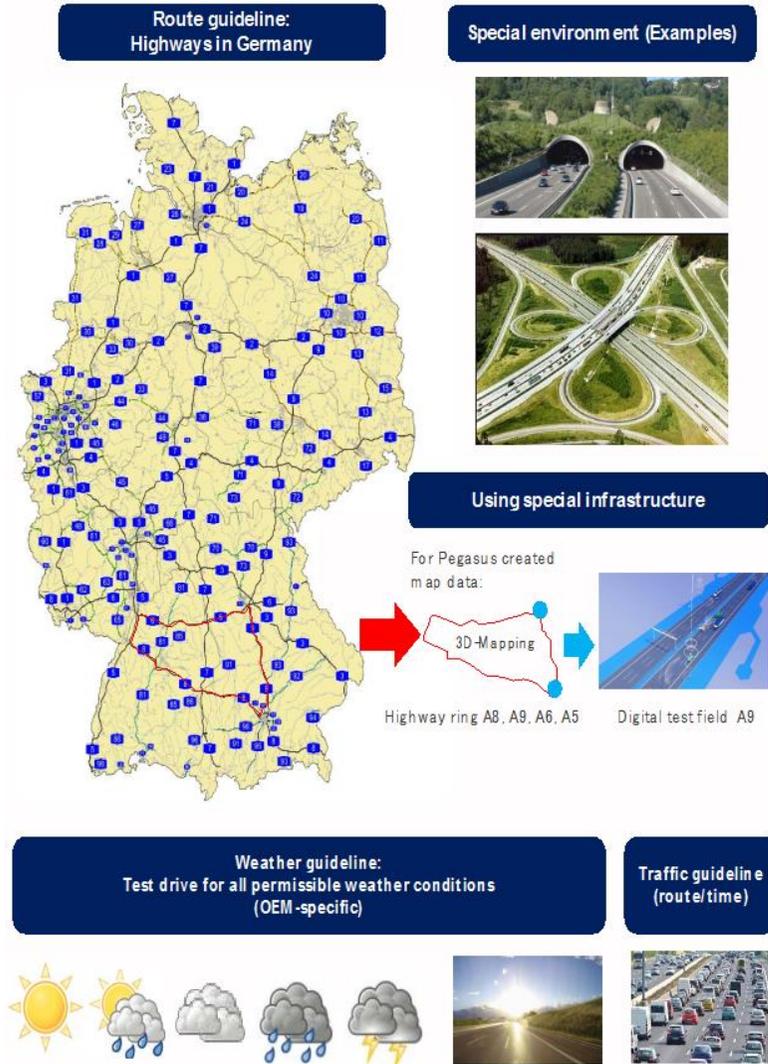
## 課題

- 自動運転車が走る物理環境の把握
- 長距離フィールド実証実験（FOT）に相当するテストの実施

## 関連する技術領域

統計\*、PEGASUS等の業界標準

\* 人間の運転と比較して安全性を主張する場合



PEGASUS Symposium 2017: 04 Scenario Description <https://www.pegasusprojekt.de/en/pegasus-symposium>

# 動向 : PEGASUSプロジェクト

## データソース



走行環境データを含むFOT



自動運転車FOT



事故データ



運転シミュレータ



テストコース走行

協調 (統合)

## PEGASUSデータベース



## テスト仕様

処理 (導出)



PEGASUS Symposium 2017: 04 Scenario Description <https://www.pegasusprojekt.de/en/pegasus-symposium>

PEGASUSは、自動運転を販売するために必要な品質基準 / ツール / 手法 / シナリオを準備

# システムにおける機械学習システム品質保証の課題

システム

→ **トップダウン**で設計するシステムと、**ボトムアップ**で訓練する機械学習モデルの**境界**

## 課題

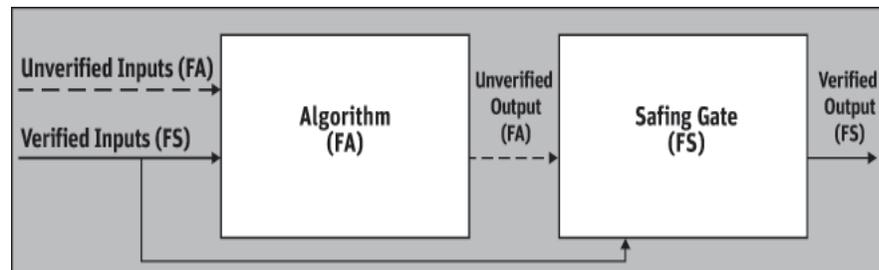
- 機械学習モデルの不確実性への対処
- 機械学習モデルのCACE性※への対処  
※Change Anything Change Everything

## 関連する技術領域

ソフトウェア工学、システム工学

## 動向

DOER-CHECKERアーキテクチャ



Source: Carnegie Mellon University

The algorithm (the “DOER”) can fail arbitrarily (FA) meaning it can do wrong things in the worst possible way.

The safing gate (the “CHECKER”) turns the algorithm into a fail silent (FS) component, only producing correct data or shutting down.

DOER-CHECKERアーキテクチャ

<https://www.ansys.com/about-ansys/advantage-magazine/volume-xii-issue-1-2018/drive-safely>

# 機械学習モデルにおける機械学習システム品質保証の課題

機械学習モデル

→ 訓練データから**ボトムアップ**で訓練（獲得）



エンジニアが設計できる最後のステップとして、機械学習モデルの要件定義が重要。

## 課題

- テストデータ分布の事前設計、事後計測
- テストデータ分布と運用データ分布の一貫性確認
- テストデータ収集の完了条件
- 多角的な品質基準（精度、頑健性、など）
- 再現可能でトレサブルな訓練
- 機械学習モデルの故障モードの定義
- 機械学習モデルに対応した自動検証手法

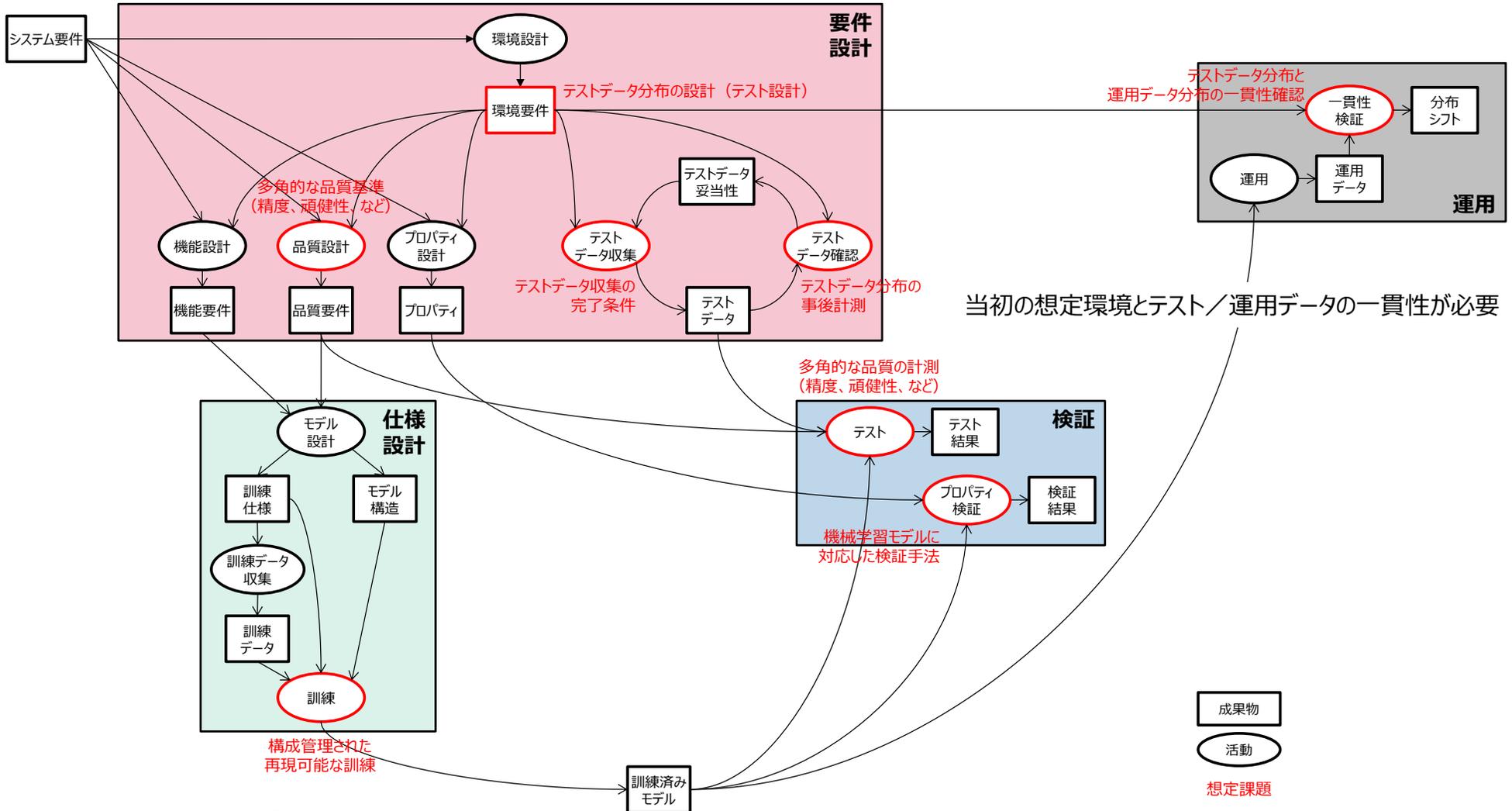
次項以降で、機械学習モデルの設計検証プロセスとともに示す。

## 関連する技術上位機

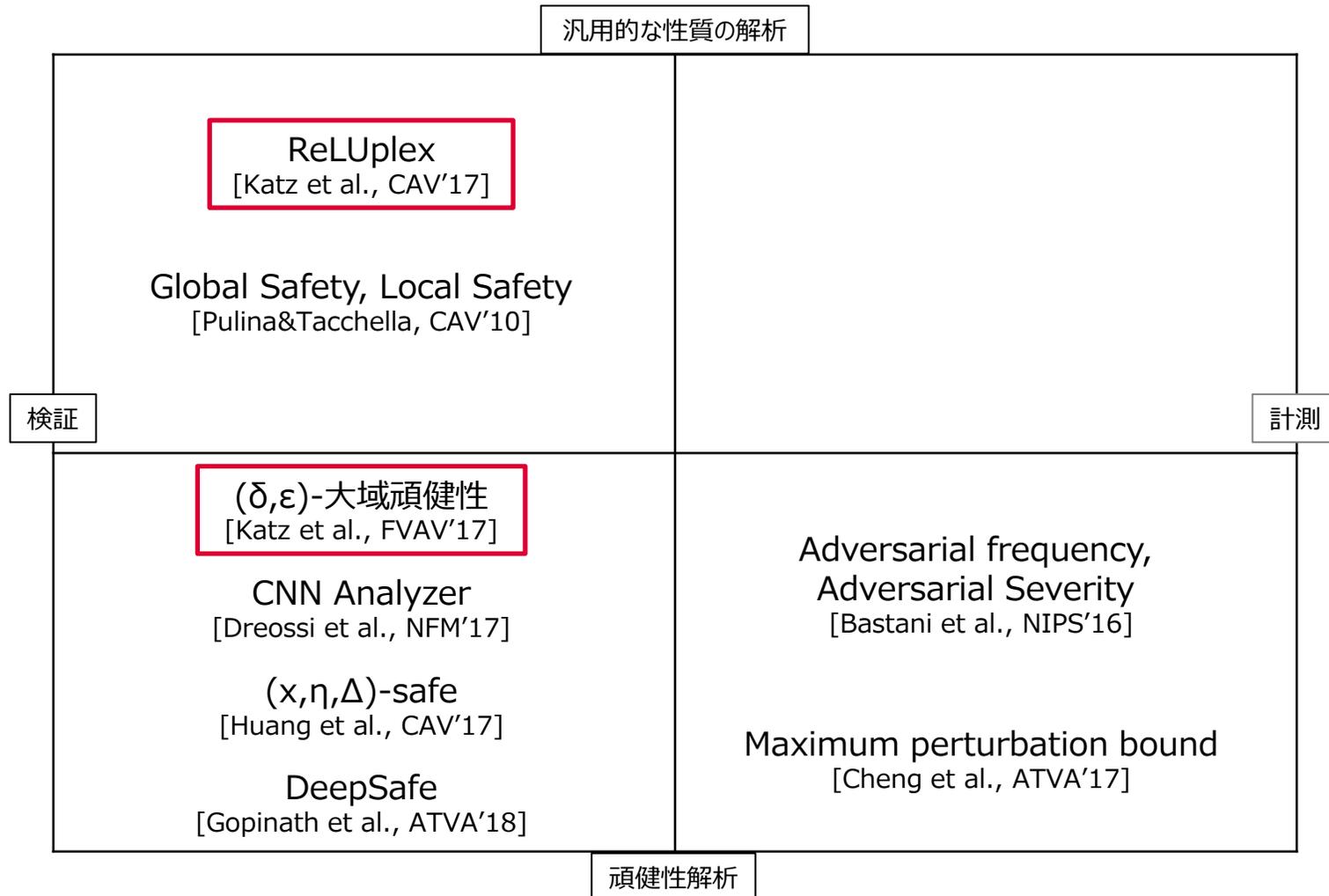
機械学習、理論計算機科学

# 機械学習モデルの設計検証運用プロセス

想定環境を定義し、それに適したテストデータ収集が必要



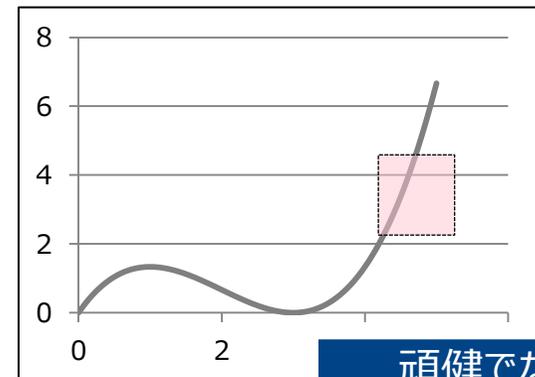
# 機械学習モデルに対応した自動検証技術



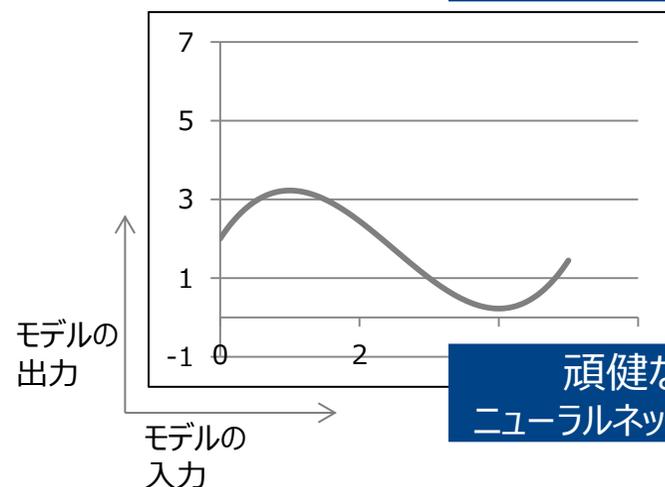
いくつもの技術が利用可能

# 機械学習モデルに対応した検証技術の例

**ReLUplex** [Katz et al., CAV'17]、**大域頑健性** [Katz et al., FVAV'17] :



頑健でない  
ニューラルネットワーク



頑健な  
ニューラルネットワーク

**大域頑健性 :**

少しの入力の変化では、大きく出力が変化しない。

**ReLUplex**は機械学習モデルの頑健性を単体法を応用して検証する。

実験例 : 下記モデルに対する検証時間  $\approx$  11時間!

- 入力数:5, 出力数:5, 隠れ層の数:6, ReLUの数:300

**ニューラルネットの頑健性を網羅的に検証するアプローチ (スケーラビリティが課題)**

# まとめ

## 背景

- ・AI～機械学習～ディープラーニング
- ・ブラックボックス化、検証手法の不在

## 動向

- ・人工知能と検証のトップ学会の発表動向
- ・機械学習に対するソフト／システム工学の学会発足動向
- ・MLSE2018（機械学習工学研究会ワークショップ）
- ・DARPA XAI（説明可能な人工知能）

## デンソーのAI品質基盤が考える課題

- ・車両、システム、機械学習モデルにおける品質保証の課題
- ・機械学習モデルに対応した自動検証技術

**セーフクリティカルな目的に機械学習システムを使うためには、品質保証の枠組みの確立が必要です  
SWEST参加者の皆様には、組み込み開発の観点から得られる課題・ソリューションの提案を期待します！**

***DENSO***

Crafting the Core