

# 深層学習の社会へのインパクト

4/7, 2017

Preferred Networks, 丸山 宏

# 汎用人工知能と特化型人工知能

-- Wired Magazineのオバマ大統領と伊藤穰一の対談 --

“There’s a distinction, which is probably familiar to a lot of your readers, between **generalized AI** and **specialized AI**. ”

汎用人工  
知能はSF  
の世界



<https://www.wired.com/2016/10/president-obama-mit-joi-ito-interview/>

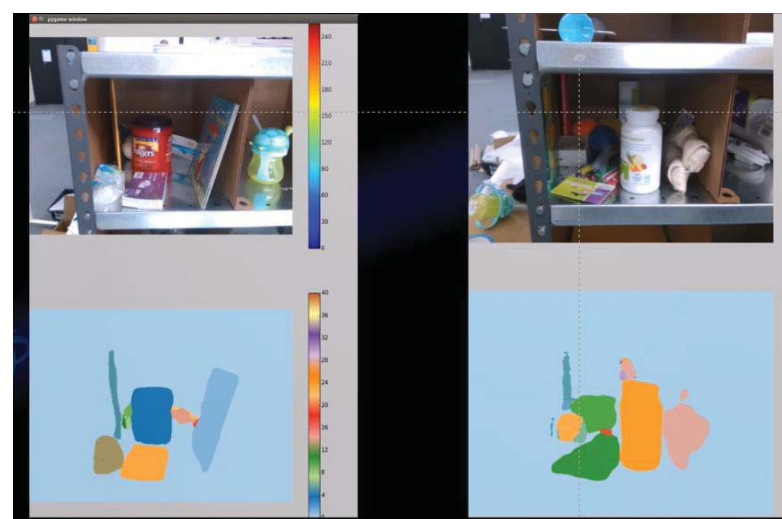
# 深層学習が可能にする世界(1) : 認識



自動運転用画像セグメンテーション



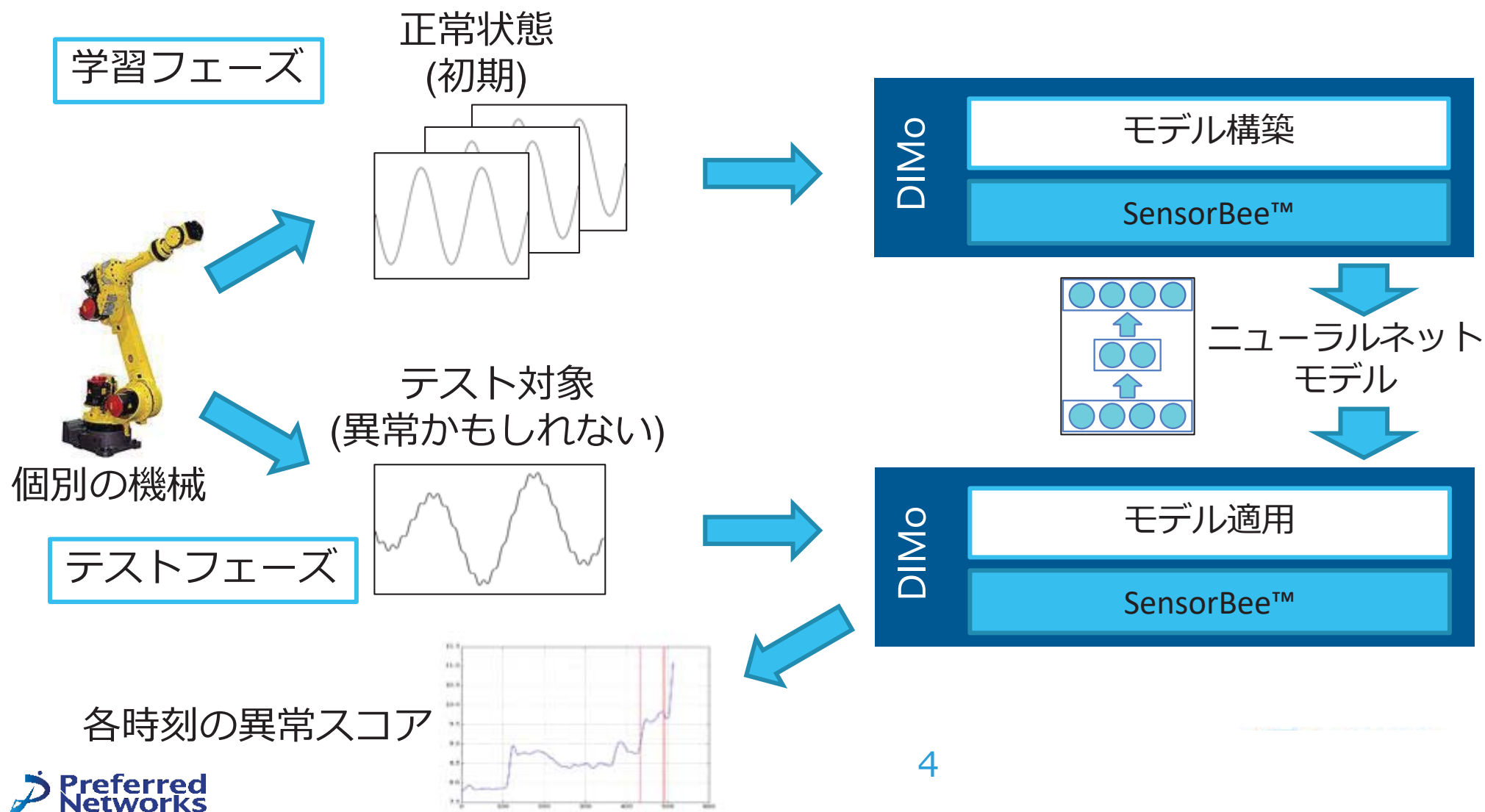
自動運転用物体認識



Amazon Picking Challenge 2016 (首位と同スコアの2位)

## 深層学習が可能にする世界(2)：予測

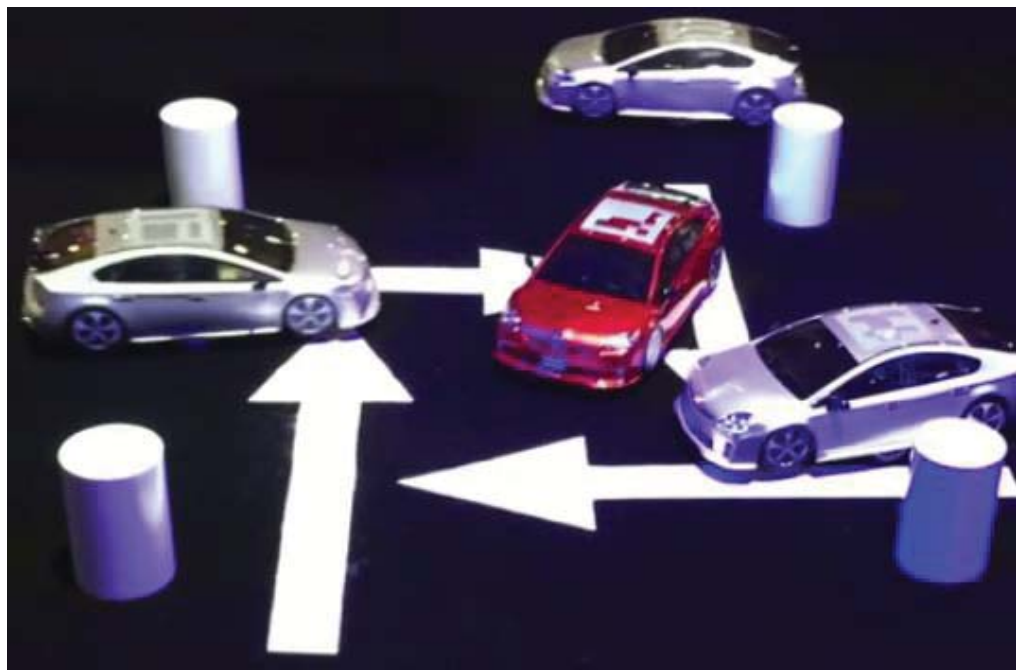
- 深層生成モデルを用いた機械の異常検出





## 深層学習が可能にする世界(3) : 制御

- 深層強化学習による機器の制御



自律走行車 (CES 2016で展示)



自律ドローン (CEATEC 2016で展示)

# 今後の機械学習/深層学習が必要とする計算リソース

学習を1日で終わらせるのに必要な計算リソース

P:Peta  
E:Exa  
F:Flops

## 画像/ 映像認識



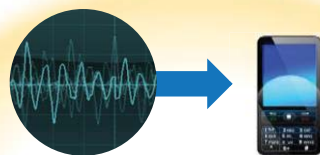
**10P (画像) ~ 10E (映像) Flops**  
学習データ: 1億枚の画像 10000クラス分類  
数千ノードで6ヶ月 [Google 2015]

## バイオ・ヘルスケア



**100P ~ 1E Flops**  
一人あたりゲノム解析で約10M個のSNPs  
100万人で100PFlops、1億人で1EFlops

## 音声認識



**10P ~ Flops**  
1万人の5000時間分の音声データ  
人工的に生成された10万時間の  
音声データを基に学習 [Baidu 2015]

## 自動運転



**1E ~ 100E Flops**  
自動運転車 1台あたり1日 1TB  
10台 ~ 1000台、100日分の走行データの学習

## ロボット/ドローン



**1E ~ 100E Flops**  
1台あたり年間1TB  
100万台 ~ 1億台から得られた  
データで学習する場合

機械学習、深層学習は学習データが大きいほど高精度になる  
現在は人が生み出したデータが対象だが、今後は機械が生み出すデータが対象となる

各種推定値は1GBの学習データに対して1日で学習するためには  
1TFlops必要だとして計算

10PF

100PF

1EF

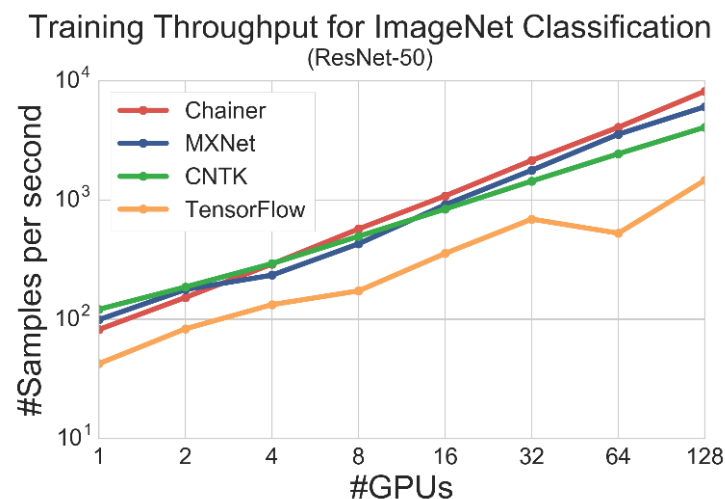
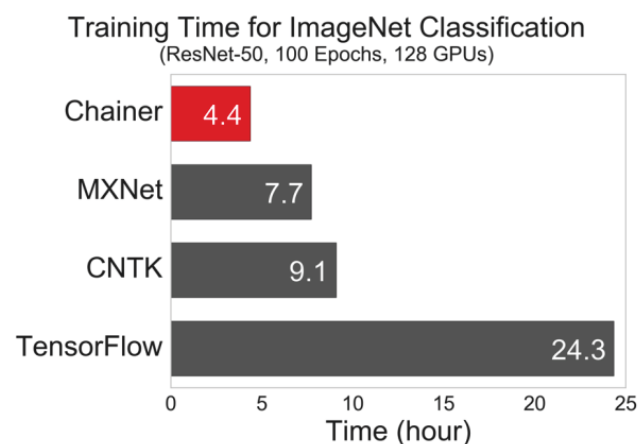
10EF

100EF

# 柔軟でスケーラブルなフレームワーク Chainer



- オープンソース
- “Define-by-run”による動的なNN生成
- 主要なプラットフォーム各社がサポート
- 高いスケール性

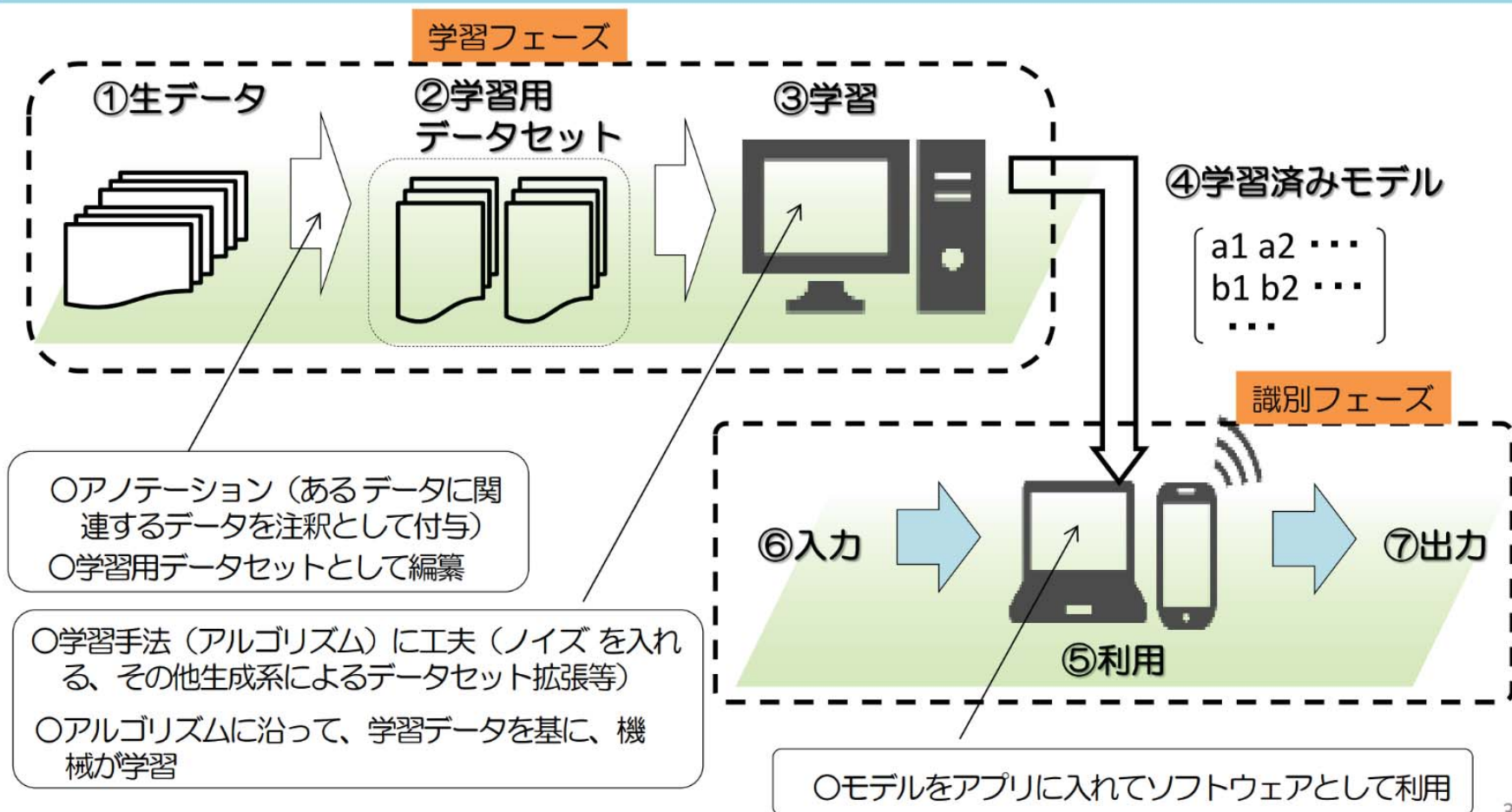


<http://chainer.org/general/2017/02/08/Performance-of-Distributed-Deep-Learning-Using-ChainerMN.html>



## 2. 人工知能（AI）（ディープラーニング）（概要）

- 人工知能は、製造プロセス、モビリティ、健康・医療、流通、インフラ・産業保安、エネルギー、行政などの幅広い分野において応用が開始されはじめている。
- ディープラーニング等の機械学習は、概ね学習フェーズと識別フェーズに分けることができ、それぞれの主な流れ（概要）は、以下のとおり。



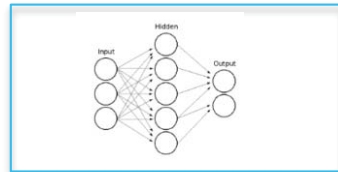
33

出典：産業構造審議会分散戦略WG



# 学習済みモデルの再利用パターン

## White box再利用



1. Copy



同タスク

2. Fine Tuning

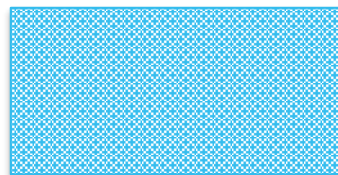


新タスク

元モデルをそのまま使う

新たな訓練データを加えて、似た領域の異なるタスクに利用する

## Black box再利用

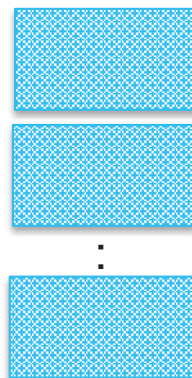


3. Distillation



同タスク

元モデルを教師として、新たなモデルを訓練する







4. Ensemble



同タスク

複数の元モデルの出力を組み合わせて精度を向上させる

# 保護に関する議論

(参考) 整理イメージ				
	特許権	著作権	不正競争防止法 (営業秘密)	一般 不法行為
①データ 	×	△ (著作物性が認められる場合は保護されるが、生データそれ自体は通常創作性が認められない。)	○ (①秘密管理性、②有用性、③非公知性の三要件を満たす場合)	×
②学習用データセット 	×	○ (情報の選択又は体系的な構成によって創作性を有するものはデータベースの著作物として保護される(著作権法12条の2))	○ (上記三要件を満たす場合)	×
③学習 	○ (特許法上の「プログラム等」に該当する場合、コンピュータ・ソフトウェア関連発明として保護される)	○ ※プログラムそのものを保護 ※リバースエンジニアリングによって同一のものが作成された場合は著作権侵害を問えない	○ (上記三要件を満たす場合) ※著作権と同様にリバースエンジニアリングには対応不可	×
④学習済みモデル $\begin{bmatrix} a1 & a2 & \dots \\ b1 & b2 & \dots \\ \dots \end{bmatrix}$	△ (プログラムに準ずるもの(「モデル」がコンピュータによる情報処理を規定するもの)に該当する場合は保護対象となるが、通常、「関数自体、行列自体」には発明成立性が認められない)	△ ※学習済みモデルが「データベースの著作物」もしくは「プログラム著作物」として著作物と認められる場合があるかは不透明 ※リバースエンジニアリングには対応不可	○ (上記三要件を満たす場合) ※頒布する場合、非公知性を満たすためには、暗号化等が必要 ※リバースエンジニアリングには対応不可	×
⑤利用 	○ (アプリ等のソフトウェアやシステムはコンピュータソフトウェア関連発明として保護される)	○ (著作物性が認められる場合) ※リバースエンジニアリングには対応不可	○ (上記三要件を満たす場合) ※頒布する場合、非公知性を満たすためには、暗号化等が必要 ※リバースエンジニアリングには対応不可	×

40

出典：産業構造審議会分散戦略WG

[http://www.meti.go.jp/committee/sankoushin/shojo/johoikeizai/bunsan\\_senryaku\\_wg/pdf/006\\_02\\_00.pdf](http://www.meti.go.jp/committee/sankoushin/shojo/johoikeizai/bunsan_senryaku_wg/pdf/006_02_00.pdf)

# 再利用に関する議論

• [Submission Guidelines](#) • [List of Topics](#) • [Committees](#) • [Invited Speakers](#) • [Venue and Date](#) • [Contact](#) • [Sponsors](#)

## SRAI\_2017: The First International Workshop on Sharing and Reuse of AI Work Products

August 19-21, 2017

### Submission deadline

May 1, 2017

As more AI technologies are used in industry applications, we anticipate that work products of AI will be shared and reused widely. Building AI-based systems involves creating and generating different types of “AI work products” such as training data sets, pretrained models, and AI-generated arts such as paintings and music. These work products are often results of heavy investment of human, data, and computer resources, and should carry some form of intellectual properties. Rules and best practices of sharing and reusing these work products are not very well agreed upon today. For the healthy proliferation of AI technologies in our society we believe now is the time to start discuss these issues. For example, when Company A applies a machine learning algorithm on a training data set prepared by Company B and generates a pretrained model of recognizing objects in images, who owns this pretrained model? How should Company C who wants to reuse this pretrained model by fine-tune it to their own domain compensate to the owner of the original model? These questions are related to various technical, legal, political, and practical issues.

This workshop is intended to be the first of series of discussions on the issues of sharing and reuse of AI work products.

### Submission Guidelines

Papers must be formatted according to the IJCAI 2017 guidelines (<http://ijcai-17.org/FormattingGuidelinesIJCAI-17.zip>), and be at most 8 pages long, including a bibliography page.

2017/8, 国際人工知能会議(IJCAI2017)で国際ワークショップ開催

Thank You