

米国 科学技術政策局における 取り組み

人工知能と人間社会に関する懇談会 事務局

概要

- 組織
 - 米国政府（科学技術政策局）
- 時期
 - 2016年から
- 内容
 - 2016年5月3日、ホワイトハウスのウェブサイト科学技術政策局次長が発表。
 - AIがもたらす利益とリスクを深く知るための活動を開始。
 - アカデミア、NPOとワークショップを共催し、AIと機械学習についての議論を促し、それらの課題とチャンス进行明らかにする。
 1. 5月24日: AIと法・統治の関わり
 2. 6月7日: 社会的利益のためのAI
 3. 6月27-28日: AIの安全性と制御
 4. 7月7日: 近未来のAI技術の社会的・経済的意味合い
 - AIと機械学習を活用して行政サービスの提供を改善するため、政府機関横断型のワーキンググループを設置。
- <https://www.whitehouse.gov/blog/2016/05/03/preparing-future-artificial-intelligence>

本日ご紹介

第3回ワークショップ：AIの安全性と制御

1日目：AIの安全性と制御の技術についての予備的ワークショップ

• 背景

- Bill Scherlis (カーネギーメロン大学)
- Ed Felten (米国 科学技術政策局)
- Eric Horvitz (マイクロソフト)
- 議論

• 展望

- Andrew Moore (カーネギーメロン大学)
- Richard Mallah (Future of Life Institute)
- Tom Mitchell (カーネギーメロン大学)
- 議論

• 焦点

- Dario Amodei (Google Brain)
- Claire Le Goues (カーネギーメロン大学)
- Robert Rahmer (米国 情報先端研究プロジェクト活動)
- 議論

- <http://www.cmu.edu/safartint/watch.html>

第3回ワークショップ：AIの安全性と制御

2日目：AIの安全性と制御についての公開ワークショップ

- 開会挨拶
 - Ed Felten (米国 科学技術政策局), Bill Scherlis (カーネギーメロン大学)
- AIアプリケーション、安全性要件、大きなアイデア
 - Manuela Veloso (カーネギーメロン大学)
 - John Launchbury (国防高等研究計画局)
 - Jason Matheny (情報先端研究プロジェクト活動)
 - Andy Grotto (米国 国家安全保障会議)
 - 議論
- AIアルゴリズム、内在的安全性、説明、評価
 - Claire Tomlin (カリフォルニア大学バークレー校)
 - Tom Dietterich (オレゴン州立大学)
 - Emma Brunskill (カーネギーメロン大学)
 - Michael Littman (ブラウン大学)
 - 議論
- 数理モデルと推論
 - Jeannette Wing (マイクロソフト)
 - Kathleen Fisher (タフツ大学)
 - Anupam Datta (カーネギーメロン大学)
 - Sarah Loos (グーグル)
 - 議論
- システムと安全工学
 - Mike Wagner (カーネギーメロン大学)
 - Drew Bagnell (Uber)
 - Reid Simmons (NSF)
 - Brian Murray (ZF)
 - Doug Schmidt (Vanderbilt)
 - 議論
- <http://www.cmu.edu/safartint/watch.html>

第3回ワークショップ：議論の抜粋



引用元: <https://www.youtube.com/watch?v=9xU-ARf6olc>

Tom Mitchell (カーネギーメロン大学)

- いかにして複雑なシステムを安全にするか。
- 例えばライフルの照準に画像認識AIを搭載し、人がターゲットになっている時や室内にいると判断した時に発射を拒絶できるようにすることで、ライフルをより安全にできる。
- 自動走行、原発、航空機、金融システム等の複雑なシステムのセキュリティ技術は互いに別物と考えられているが、AIに基づく汎用技術があり得るのではないか。

Dario Amodi (Google Brain)

- AIの安全性に関する議論は super-intelligence 等の極端なケースを想定したものが多いが、今その心配をするのは火星での人口過剰を心配するようなもの。現存するシステムに関連付けられる具体的、経験的な議論をしたい。
- AI (機械学習) のアクシデントの原因の1つに目的関数の間違いがある。例えば掃除ロボの目的関数を「ゴミが見えなくなればOK」としてしまうと、掃除ロボは目をつむって目的達成としてしまう。



引用元: <https://www.youtube.com/watch?v=F25i0sgrp9M>

第4回ワークショップ：近未来のAI技術の社会的・経済的意味合い

専門家向けワークショップ

- イントロダクション
- 今回は違うのか？ AIのチャンスと課題
 - Jason Furman（大統領経済諮問委員会）
- 講演 -不平等と労働-
 - 大分断 Erik Brynjolfsson (MIT)
 - マシンによる偏見を明らかにする Julia Angwin (ProPublica)
 - AIを「魔法」にする労働 Lilly Irani (カリフォルニア大学サンディエゴ校)
 - 今回は違うのか？ レース、労働、AI Dorian Warren (Roosevelt Institute)
 - 共生的人-ロボットインタラクション Dawn Gearhart (International Brotherhood of Teamsters)
 - 誰がAIについて考えるようになるのか Darius Kazemi (Feel Train)
- 講演 -健康と倫理-
 - 機械加工の倫理 Ryan Calo (ワシントン大学)
 - 人と機械の新しい関係を通して人間の条件を高める Guru Banavar (IBM Research)
 - 機械学習と医療：リスクと恩恵 Jennifer Chayes (Microsoft Research)
 - 熟慮の場 Paul Dourish (カリフォルニア大学アーバイン校)
 - 補完的 vs 代替的医療オートメーション Frank Pasquale (メリーランド大学)
 - 分類される私たち Jer Thorp (Office for Creative Research)

第4回ワークショップ：近未来のAI技術の社会的・経済的意味合い

公開シンポジウム

- インTRODクシヨN
- ホワイトハウスからの開会挨拶
- ホワイトハウスとの会話、過去と未来
 - David Edelman (大統領経済技術政策特別顧問)
 - Roy L. Austin (大統領副補佐官)
 - Nicole Wong (前 合衆国次席技術専門官)
 - Kate Crawford (司会)
- ハイテク業界リード企業3社に対する3つの質問
 - Genevieve Bell (Intel)
 - Cynthia Breazeal (Jibo)
 - Yann LeCun (Facebook)
 - Julia Angwin (司会)
- 討論：AIにおける不平等、労働、健康、倫理
 - Henry Siu (ブリティッシュコロンビア大学)
 - Lucy Suchman (ランカスター大学)
 - Mustafa Suleyman (DeepMind)
 - Latanya Sweeney (ハーバード大学)
 - Alondra Nelson (司会)

第4回ワークショップ：議論の抜粋



Darius Kazemi (Feel Train)

- AIを使うにはそのためのモデルを機械学習しなくてはならないが、学習には大規模な計算機パワーが必要。
- そうした計算機パワーを持てるのは企業、国、大学等だけであり、個人は企業等があらかじめ学習したモデルを使わざるを得ない。
- そうしたモデルには予期せぬ問題（例：黒人女性をゴリラと誤認識）がありうるが、個人にはコントロールできない。

引用元: <https://artificialintelligencenow.com/schedule/conference/presentation/who-gets-think-about-ai>

Jennifer Chayes (Microsoft Research)

- 乳癌の原因となる遺伝子変異は様々で、変異のタイプをAIで特定することが有効。
- 糖尿病患者に運動させるため、AIを使って患者個人に特化したフィードバックを与えたところ、そうしなかった患者より運動するようになった。
- こうしたAIは膨大な個人情報をもとにしており、個人にとってはリスクがある。AIがもたらす恩恵のためにどの程度リスクを許容できるか。セキュリティ技術が重要だろう。



引用元: <https://artificialintelligencenow.com/schedule/conference/presentation/machine-learning-and-healthcare-risks-and-rewards>