

人工知能（AI）と 私たちの社会

東京大学国際高等研究所東京カレッジ

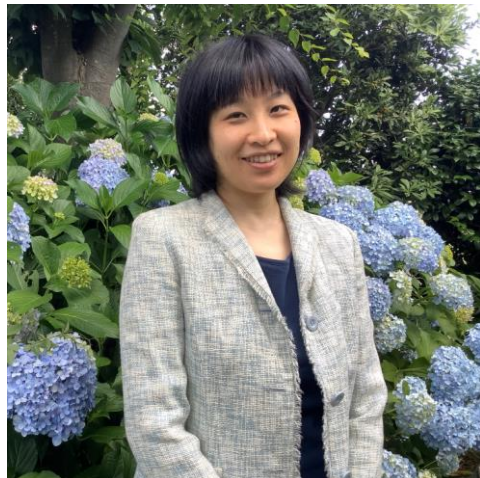
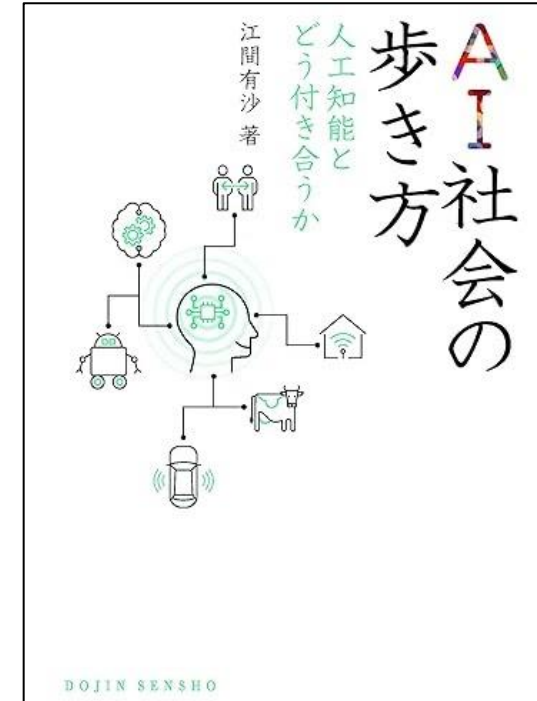
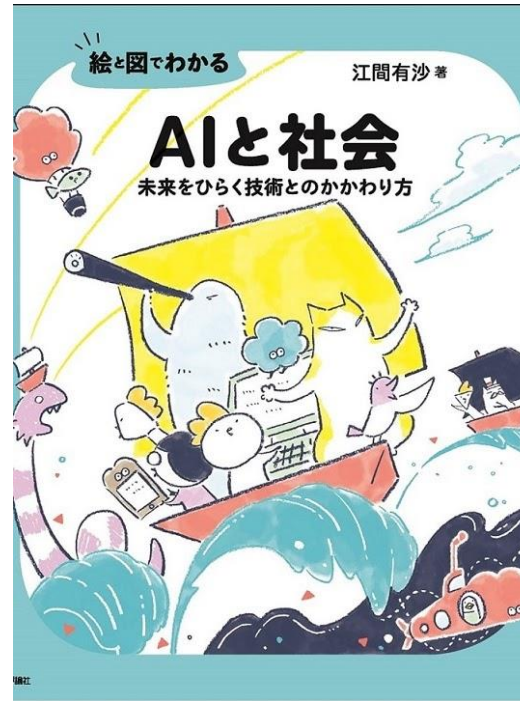
江間有沙

江間有沙 (えまありさ)

東京大学東京カレッジ 准教授
日本ディープラーニング協会 理事

キーワード：AIと社会、ガバナンス
委員など

- 東京都「AI戦略専門家会議」委員
- 文部科学省「初等中等教育段階における生成AIの利活用に関する検討会議」委員
- 神戸市AI活用アドバイザー
- 内閣府「イノベーション政策強化推進のための有識者会議『AI戦略』」構成員
- 総務省「安心・安全なメタバースの実現に関する研究会」構成員
- 内閣府「人間中心のAI社会原則会議」構成員



人工知能とは

- 1956年に開催されたダートマス会議で、アメリカの計算機科学者であるジョン・マッカーシーが提唱
- 人間の知的活動を定義し、技術的に実現する研究分野
- 大量のデータを学習することで、自ら特徴やパターンを抽出し、初めて見るデータでも分類、識別や予測ができる技術
- 人工知能技術のうち、機械学習は統計学を基礎としているため、過去と未来は変わらないという前提で認識や予測を行う
- つまり、現在の社会に存在する差別や偏見も再生産してしまう

人工知能技術と差別や偏見に関する事件簿

- 学習
 - SNSのコメントで、機械が差別的発言をする
→システム凍結
- 認識
 - 顔認識システムで、「浅黒い肌の女性」の認識
精度/正答率が低い
→警察など公的機関での利用を制限
- 評価
 - 採用判定システムが、女性に関する単語が含まれていると評価を下げる
→システム開発断念

クイズ：

このシステムは合憲？違憲？

- アメリカのある州では、過去の犯罪データと照合して、犯罪者の再犯可能性を予測するシステムが使われている
- システムの判断に基づいて、裁判官が判決を下す
- このシステムが人種差別をするとして州最高裁で争われた

再犯リスク予測AI COMPASの事例

社会的規準「公平」を数学的表現に落とし込む



報道機関

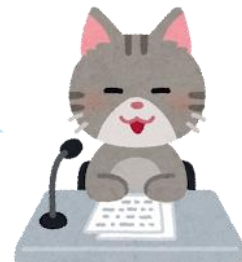
「再犯率が**高い**と予測されたが、実際には再犯しなかった率が黒人のほうが多かった。黒人のほうが「再犯する」と思われる偏見のある**不公平なシステム**である！

「再犯率が**低い**と予測されたが、実際には再犯しなかった率」が人種に寄らず一致するので**公平なシステム**である！



開発者

両方を同時に満たす公平性は存在しないので
何が「**公平**」かは社会的に決めて技術に反映させる必要がある
→様々な人たちの協力が大事になる



合憲：システムを利用してよい

ただし！

- システムは参考として使用し、総合的に考えて人間が最終判断をする
- システムの判決に人種の偏りがあると利用者に注意を促す必要がある
- 人間がAIの課題を認識しつつ、利用することが重要となる

QuizKnock と探る!



未来ビジョン

#2



東京大学未来ビジョン研究センター
Institute for Future Initiatives
The University of Tokyo

AIは私たちの社会の鏡

- You are what you eat
- しかし一方で、その鏡が歪んでいることもある
 - ステレオタイプが増産されていることもある

◆画像検索や生成をしてみよう

- 「社長」や「医者」
 - 2020年4月末における女性社長の割合は8%
 - 2018年度末地点で、医師の男女比21.9%
- 海外の検索ウェブサイトで「CEO」や「Doctor」, 「Baby」
 - .comや.ukがある検索サイト

どういう基準でシステムを設計する？

- 手続きが公平か結果が公平か
- 現実と同じであること
- 弱い立場にいる人たちに配慮すること
- 大多数が満足していること

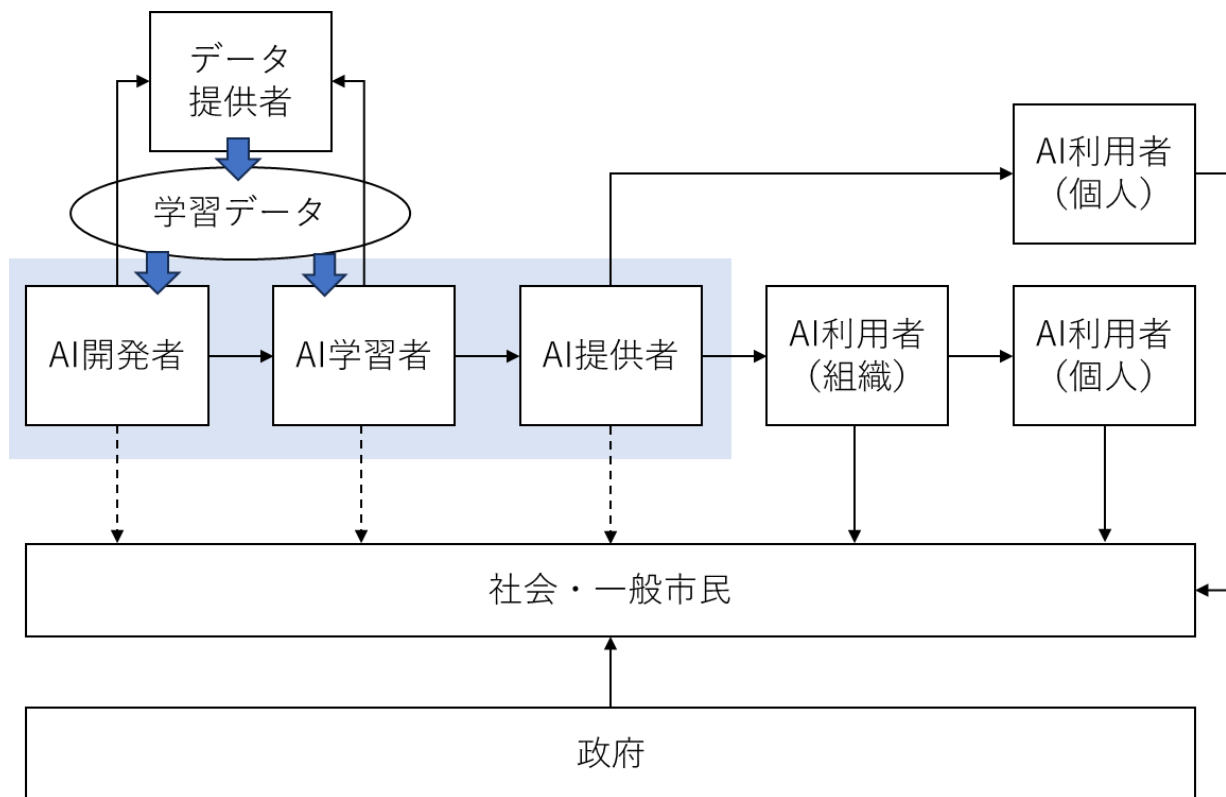
◆ 偏りを見つけたらどうする？

1. 何もしない
2. 少なくとも現状に合わせる
3. 理想に合わせる

生成AIをめぐる議論

- 「AIは正しい情報を出さないから、気をつけろ」
 - 人間であれば正しい判断ができるのか？
 - なぜ機械だと許せないのか？
- AIはそのような人間から学習する
 - 逆にAIを反面教師として「公平」な出力をすることも可能？
 - ただし何が「公平」なのかは議論する必要がある
- AIにできること、できないことを理解して使うこと

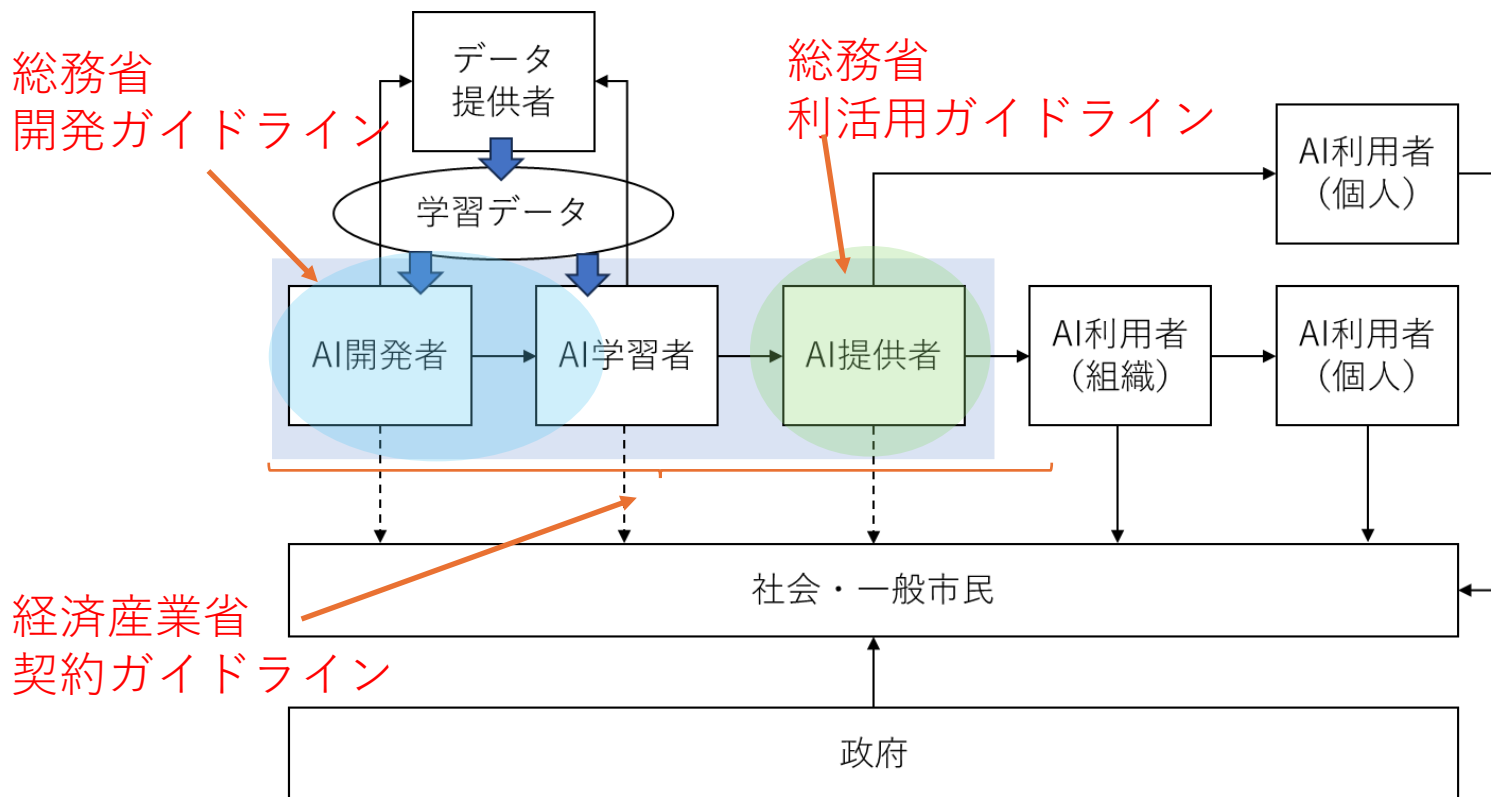
誰が誰に対して責任を持つか？



AI開発者・AI学習者・AI提供者は同一組織の場合もあれば、別々の組織の場合もある
また活用户（組織）や利用用户も同一になる場合もあるが、今回は考えない。

- 責任の向きを示す。例えばAI開発者はAI学習者とデータ提供者に対して責任を負う
- - - - -> 直接的ではないが、間接的な責任の向きを示す。例えばAI開発者は社会・一般市民に対し、研究の内容を説明したり、発信したりするという社会的な責任がある。

既存のガイドライン

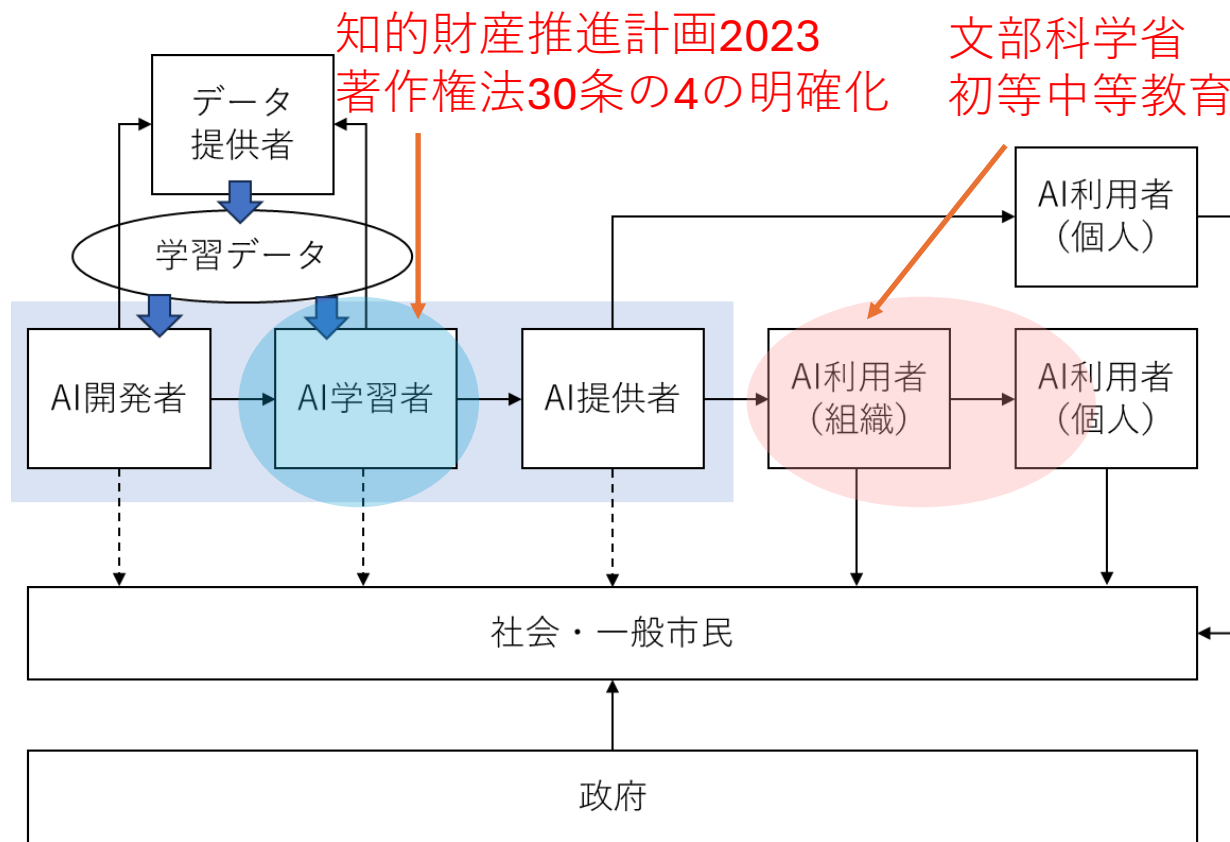


AI開発者・AI学習者・AI提供者は同一組織の場合もあれば、別々の組織の場合もある
また活王者（組織）や利用者も同一になる場合もあるが、今回は考えない。

- 責任の向きを示す。例えばAI開発者はAI学習者とデータ提供者に対して責任を負う
- - - - -> 直接的ではないが、間接的な責任の向きを示す。例えばAI開発者は社会・一般市民に対し、研究の内容を説明したり、発信したりするという社会的な責任がある。

生成AIの焦点

日本ディープラーニング協会
生成AIの利用ガイドライン



AI開発者・AI学習者・AI提供者は同一組織の場合もあれば、別々の組織の場合もある
また活字者（組織）や利用者も同一になる場合もあるが、今回は考えない。

- > 責任の向きを示す。例えばAI開発者はAI学習者とデータ提供者に対して責任を負う
- - - - -> 直接的ではないが、間接的な責任の向きを示す。例えばAI開発者は社会・一般市民に対し、研究の内容を説明したり、発信したりするという社会的な責任がある。

私たちはどのような社会に
住みたいのか

人間が想像できることは、
人間が必ず実現できる

ジュール・ヴェルヌ
(仏小説家：『海底二万里』など)

科学技術と社会：未来を考える

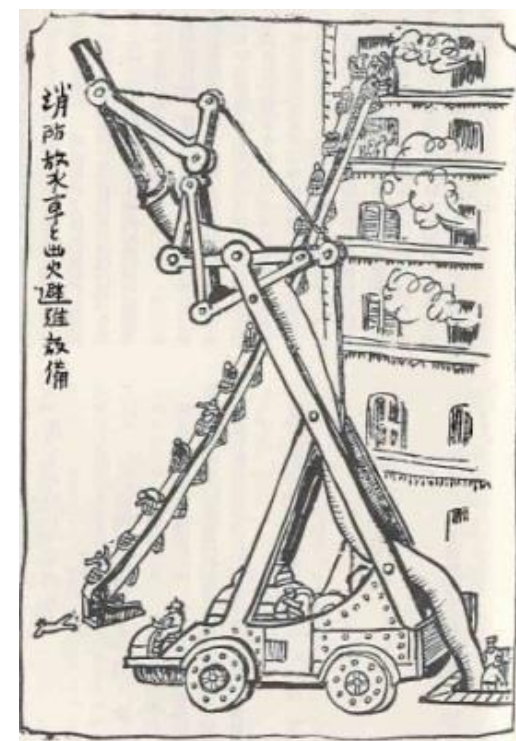
- 科学技術の社会面
- 社会の中の科学技術
 - どういう社会になっていくか
 - どういう未来に私たちは住みたいのか



Jean-Marc Cote

“In the Year 2000”

～1900年パリ万博展示ポストカード～



三宅雪嶺主催『日本及び日本人』
百年後の日本特集（1920）

例) 目指すべきとされる未来社会

- 健康：健康寿命が長く、病気も予防される
- 幸福：自分の人生や仕事に生きがいを持つ
- 安定：人々も社会も不安定になることがない
- 繋がり：孤独になることはない、だれも取り残されることはない

- 隠し事もなく、嫉妬もなく、誰もが他のみんなのために働いている社会

60歳ぐらいで死ぬまで、ずっと老
いずに若い。

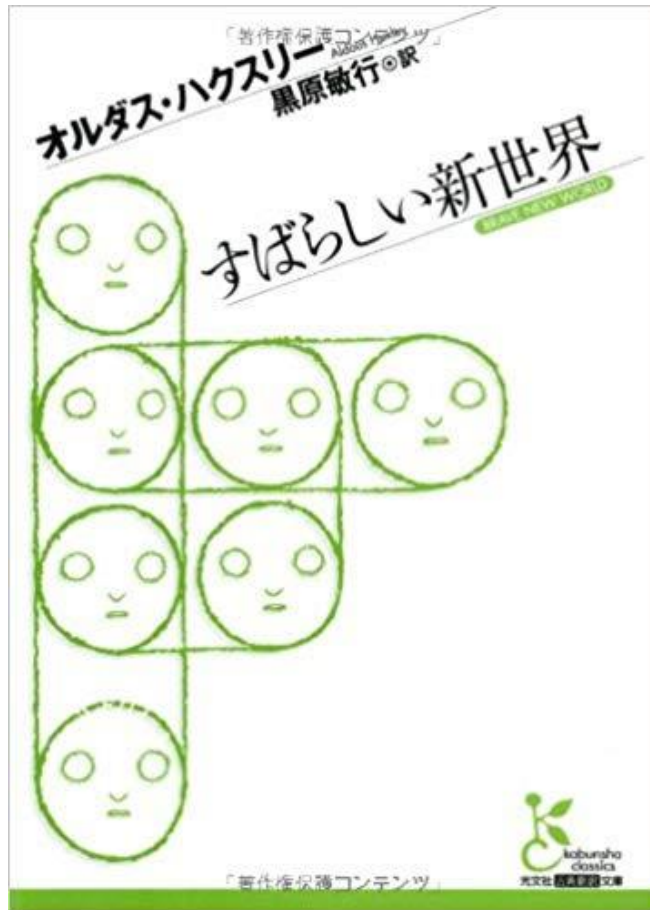
人々は生活に完全に満足してい
る。

人々は、激情に駆
られることなく、常に安定した精神状態である。
そのため、社会は完全に安定している。

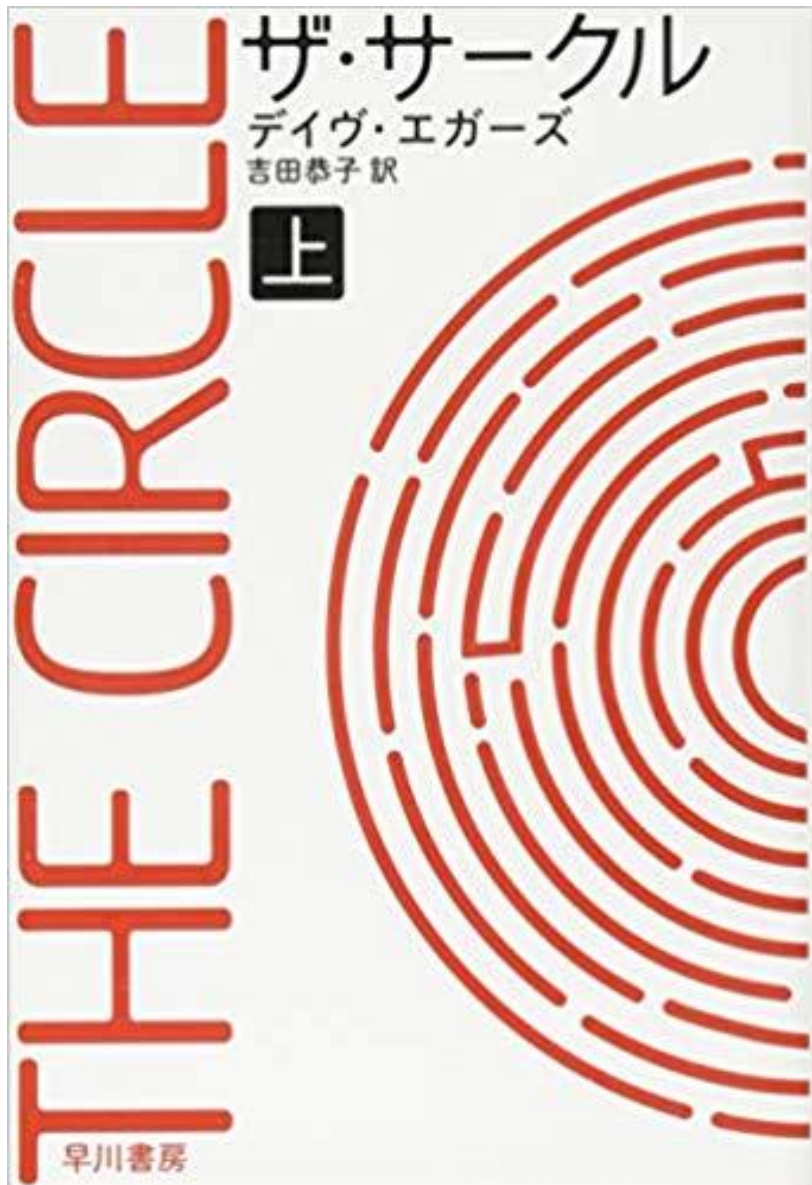
人々は常に一緒に過ごして孤独を感じるこ
とはない。

- 隠し事もなく、嫉妬もなく、誰もが他のみんなの
ために働いている。

ユートピア/ディストピア小説



- また、あらゆる予防接種を受けているため病気になる事は無く、**60歳ぐらいで死ぬまで、ずっと老いずに若い。**
- ビンから出て「出生」した後も、**睡眠時教育で自らの「階級」と「環境」に全く疑問を持たないように教え込まれ、人々は生活に完全に満足している。**
- 不快な気分になったときは、**「ソーマ」と呼ばれる薬で「楽しい気分」になる。人々は、激情に駆られることなく、常に安定した精神状態である。そのため、社会は完全に安定している。**
- **ビンから出てくるので、家族はなく、結婚は否定されて人々は常に一緒に過ごして孤独を感じることはない。**
- **隠し事もなく、嫉妬もなく、誰もが他のみんなのために働いている。一見したところではまさに楽園であり、「すばらしい世界」である。**

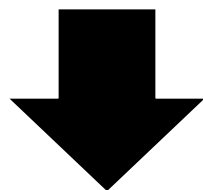


- ザ・サークル：エマ・ワトソン×トム・ハンクス主演映画化!
- 世界一と評されるインターネット企業、サークル。
- 広々とした**明るいキャンパス**、
- 信じられないほど**充実した福利厚生**
- **頭脳と熱意と才能をかねそなえた社員たち**が次々に生み出す新技術——そこにはないものはない。
- サークルへの転職に成功した24歳のメイは、新生活への期待で胸をいっぱいにして働きはじめるが…
- SNSとウェブの未来を予言するサスペンス
- サークル社の社是は、**"秘密は嘘。分かち合いは思いやり。プライバシーは盗み"**

The road to hell is paved with good intentions
地獄への道は善意で舗装されている



- コリングリッジのジレンマ
 - 技術が社会で使われる前にその影響力を予測することは難しい
 - 一度普及してしまった技術は制御するのが難しい



- 「社会実験」ではなく「実験社会」
- 私たち一人一人の役割と責任とは？
- どのような生き方・働き方を目指す？

AI社会の歩き方

- 技術を道具として使う
 - 特性を理解して賢く使うこと
- 人間の方が場合によってはブラックボックス
 - 社会そのものが偏っていることも
 - 機械の方が「人間の思い込みや私情」が入らない場合もある
- 未来を作るのは人間と機械の協同作業
 - 鏡としてのAIから、社会の歪みが見えることもある。過去からの学び
 - どういう社会に住みたいか？に答えられるのは人間。未来への創造力

常識とは十八歳までに身につけた偏見のコレクション

アルベルト・アインシュタイン (1879-1955)

言語化する方法論

タスク分けワークショップ

ワークショップの目的

- 人工知能は面白い
 - ビジネスモデル・政策と相互作用して知らないうちに生活に入り込んでいる
 - 国内外で議論が盛ん
- でも、それに「あなた」がどう感じるか？
 - ◆わくわくする：自分の価値と合う
 - ◆もやもやする：自分の価値に挑戦する
 - ◆いらいらする：自分の価値に反する

不安や期待を共有してみよう

ワークショップのルール

1. 自分の思いを発してみる
 - 自分事として意見をぶつけてみよう
2. 批判はしない
 - 「議論」ではなく「対話」をする
 - 一人一人の考えや受け止め方を大切にしよう
3. 気になったら、聞いてみる
 - お互いの違いや多様性から学んでみよう
4. 話題を変えることを恐れない
 - いろんな視点から議論してみよう
5. フラットな関係で議論しよう
 - お互いを「さん」づけで呼ぼう

意見はポストイットに
書いて共有してみよう。
他の人の意見で面白
かったものも、メモし
て残しておこう

自己紹介（一人15秒！）

- 名札にポストイット、「今日のニックネーム」
 - 読みやすい、呼びやすい、他の人と被らない
- 自己紹介方法
 - ×「△△研究科の**です」
 - ○「（ニックネーム）です」
- 「人工知能」に対する印象を一言
 - 例「もやもやしてます」「ディープラーニングが気になります」「何が人工知能なのかわかりません」「医療応用に期待してます」 etc

ワークショップのテーマ

- そもそも自分の「暮らし」を見直してみよう
 - 科学技術など関係なく
 - 今日、朝起きてからやったこと
 - 今、何に困っているか、何が楽しいか
 - 「タスク」単位で書き出してみよう
 - ゴミ出し、ペットの餌やり、後輩の教育、メールの返信、栄養のある食事作り、友達へのメールの返信

用意されたタスク

- 要介護者の排泄処理
- 楽器の演奏
- ペットの餌やり
- 論文を読む
- 服を選ぶ
- ごはんを作る
- 友人との日常会話
- 指導教員との日常会話
- 愚痴を聞いてもらう
- 進路を決める
- 選挙で投票する
- 裁判で判決・量刑を下す

まずは用意されたタスクを書いて貼ってみよう
(タスクをさらに分解するのもOK)

ポストイットの整理

機械に任せたいタスク
(合意有)

合意が得られな
かったタスク

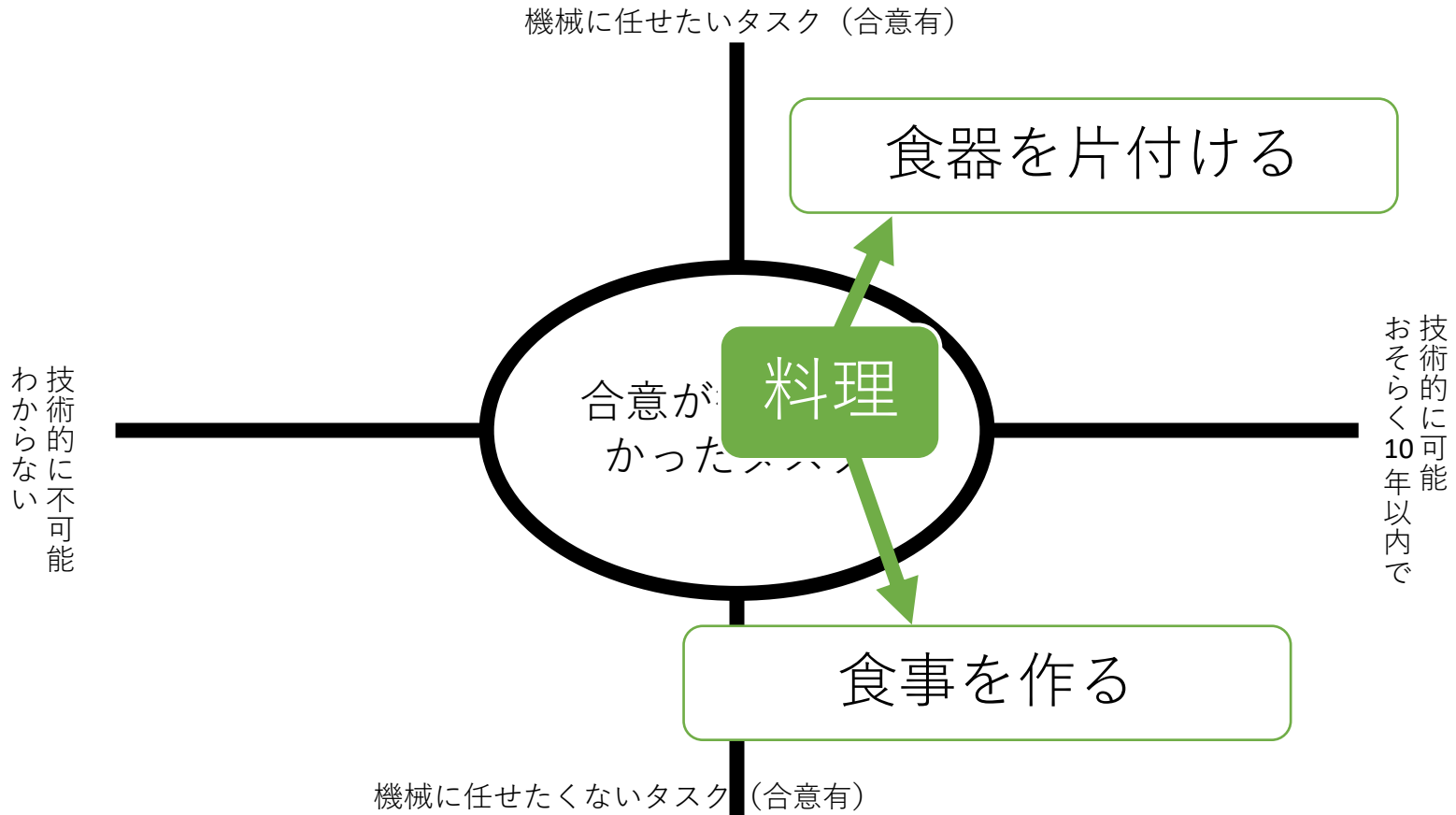
機械に任せたくないタスク
(合意有)

技術的に不可能
わからない

技術的に可能
おそらく10年以内で

ポストイットを貼ってみよう

- 他の人のアイデアに乗っかって書いていこう
- 合意ができなかったら分割してみよう



機械に任せたい

介護

10
年
以
内
に
可
能

10
年
以
内
に
不
可
能
わ
か
ら
な
い

選挙

機械に任せたくない

全体共有（10分）

- 各テーブル1分（写真）
- ファシリテーターの独断と偏見で発表
 - 全体的な印象（4象限のうちタスクの出やすさ等）
 - 面白かったタスクは何ですか
 - 合意をとるのにもめたカードはありますか？
 - タスクを分けたことで合意が取れたものはありますか？
 - 席替えした時の印象 等

任せたい

悪癖を聞いて
もらう

ノット
開閉

事件簿の
判事処理

筆器の渾身

合意で進めた

指導教員の
日常会話

版を
内には
能

選挙で
投票

論文を
読む

裁判で
判決・量刑

不明

これを
作る。

友との
日常会話

任せたくない

進路も決める

任せたら...

ゼミの
司会進行

バイト
(休)

要介護者の
排泄処理

魚のえさやり
②'

KY ③

適合度 ④'

内容把握
④'

不可 in 10年

① 進路を
決めよう
論文を読む
④
ハットの
えさやり
②

裁判で判決
量刑

指導教員との
日常会話

10年でOK

進路の判断
①' 決定

研究についての
会話
(指導教員との)

服も選ぶ

インスピレーション
④' 得る

KYX ①'

ご飯を作る

74を
聞いてもらう

友だちとだべる
③
犬猫のえさやり
②'

楽器演奏

任せたくない

表3 大学院生・人工知能学会参加者と中学生のタスク比較表

	大学院生・人工知能学会参加者	中学生
意思決定系タスク 何かを決定したり判断したりする 内容が入っているタスクの系統	選挙で投票する	生徒会選挙で投票する
		AKB選挙で投票する
	進路を決める	進路を決める
	服を選ぶ	服を選ぶ
	裁判で判決・量刑を下す	—
会話系タスク 人間関係に関わるタスクの系統	要介護者の排泄処理	お年寄りのお世話
	友人との日常会話	友人との会話
	指導教員との日常会話	先生との会話
	愚痴を聞いてもらう	悩みを聞いてもらう
勉強系タスク 研究、勉強に関わるタスクの系統	論文を読む	教科書を読む
		勉強する
日常系タスク 日常、趣味に関わるタスクの系統	楽器の演奏	楽器の演奏
	ペットの餌やり	ペットの餌やり
	ご飯を作る	—

生成AI社会の生き方

- 技術を道具として使う
 - 特性を理解して賢く使うこと
- 人間の方が場合によってはブラックボックス
 - 社会そのものが偏っていることも
 - 機械の方が「人間の思い込みや私情」が入らない場合もある
 - 人によって機械に任せたい・自分でやりたいことは異なる
 - 対話をしていくことこそが重要
 - そのための場の一つが学校
- 未来を作るのは人間と機械の協同作業