人間中心のAI社会とは
未来への責任を考える

直江清隆(東北大学)

未来への倫理

価値発見的な視点とは

- 今後は、人文・社会科学の「知」と自然科学の「知」を融合した「総合知」がますます重要となる。言い換えれば、人文・社会科学の真価である**価値発見的（heuristic）な視座**を、科学技術・イノベーション政策と融合していくことが求められるだろう

総合科学技術・イノベーション会議 基本計画専門調査会
『科学技術・イノベーション基本計画の検討の方向性（案）』（2020年）、6頁

価値発見的な視点とは

- 新たな社会を設計し、その社会で新たな**価値創造**を進めていくためには、多様な「知」が必要である。特にSociety 5.0への移行において、新たな技術を社会で活用するにあたり生じる**ＥＬＳＩ**に対応するためには、俯瞰的な視野で物事を捉える必要があり、自然科学のみならず、人文・社会科学も含めた「総合知」を活用できる仕組みの構築が求められている

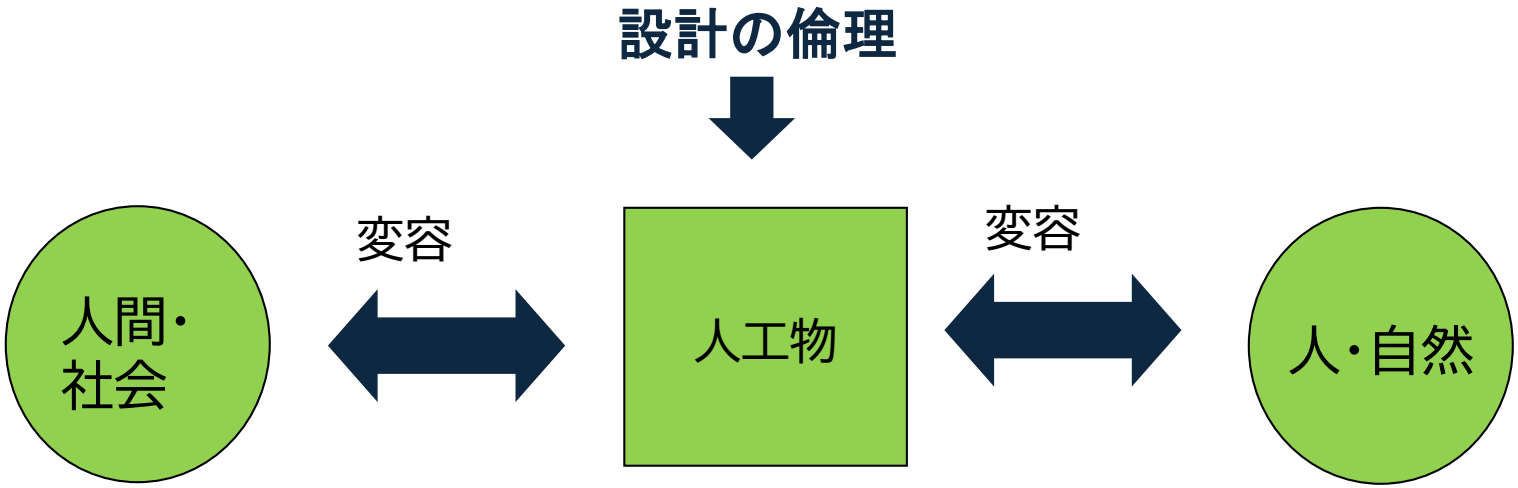
『第6期科学技術・イノベーション基本計画』（2021年）、14頁

- **技術者倫理**
- 専門的知識、自律性、社会的責任をもつ専門職としての技術者がもつべき倫理（公衆の安全、誠実さや客観性、利益相反の回避、守秘義務……）〈**ひとの倫理**〉

- **設計の倫理**
- 設計者は人工物を設計することでもう一つの仕方で倫理に携わっている。（人工物の政治性や倫理性）〈**ものを介した倫理**〉

- 人工物がないとき

- 人工物が介在するとき



ネガティブな倫理

- 何かを制限し、私たちが何をしてはならないかを言う倫理

ポジティブな倫理（Positive ethics）

- 善き生と善き社会のビジョンに基づいて、私たちが（共に）いかに生きるかに関係する倫理。前向きの倫理の一種

ポジティブな倫理　補足

前向き責任 (forward-looking responsibility)　⇔　後向き責任・非難責任
いま起こりつつある出来事やこれから起きる出来事に関して行為者が負う責任。後ろ向きの帰責と対比。これから起きる出来事に対する予防策を含む。

志向倫理(aspirational ethics)　エンジニアに対したんなる義務（「やってはならないこと」）を超えて人間のwell-beingを促進することを要求(Harris, 2017)。「誇り」「社会意識」環境意識」といった「プロフェッションとしての品性」として「誇り」「社会意識」環境意識」など。予防策と対立しうる。

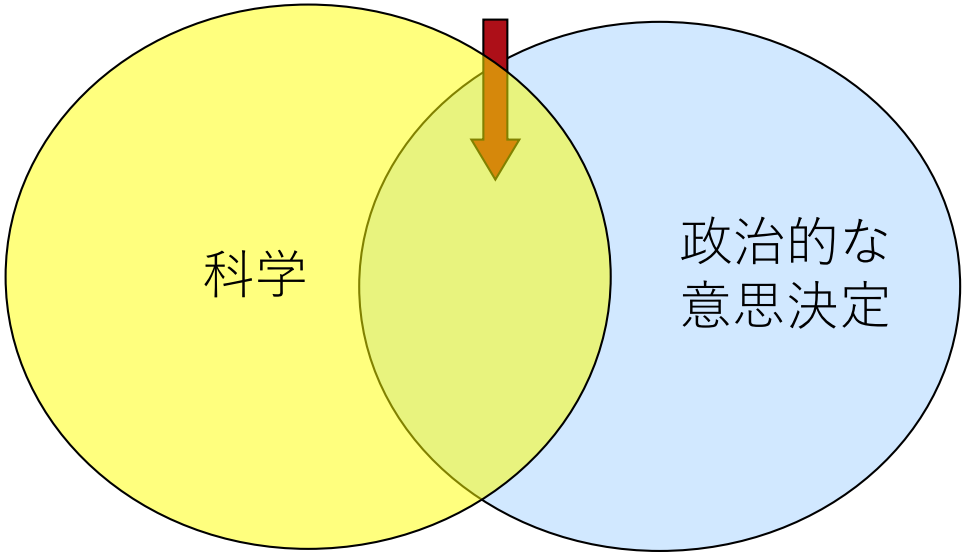
ポジティブな倫理
善き生活と善き社会のビジョンに基づいて、私たちが（ともに）いかに生きるかに関係する倫理。　これは、制限を設けて、私たちがしてはならないことを述べる否定的な(negative)倫理とは対照的。「価値に配慮したデザイン」(Value-sensitive Design)を含む。

人間中心的ということ

科学と技術における意思決定

「トランス・サイエンス的問題」(Weinberg)

科学の言葉で記述することはできるが、科学によって（科学的な定量的評価によって）答えることができない問題



Technological Fix （技術的問題解決）

もっぱら技術的手段によって社会問題の解決を図る

例えば、核兵器を戦争抑止力とする

学級崩壊にあるクラスの生徒に精神安定剤を用いる

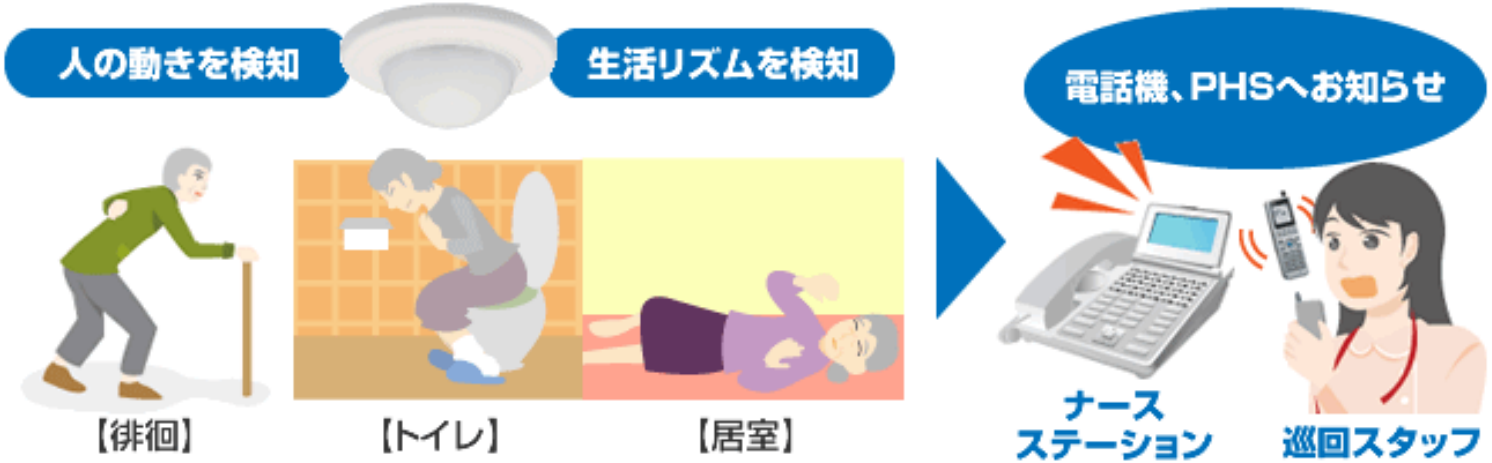
Technological Fixでは、問題と解決が両立しない

「テクノロジカル・フィックスは、問題の核心に切り込むことはしない。それはいっときの応急策にすぎず、もとの問題の代わりに新たな問題をつくり出す」（ワインバーグ）

「技術が拓く未来」から
「技術によって拓く」人間中心社会への転換

例：介護用ロボット

人感センサー(看護・介護用センサー)



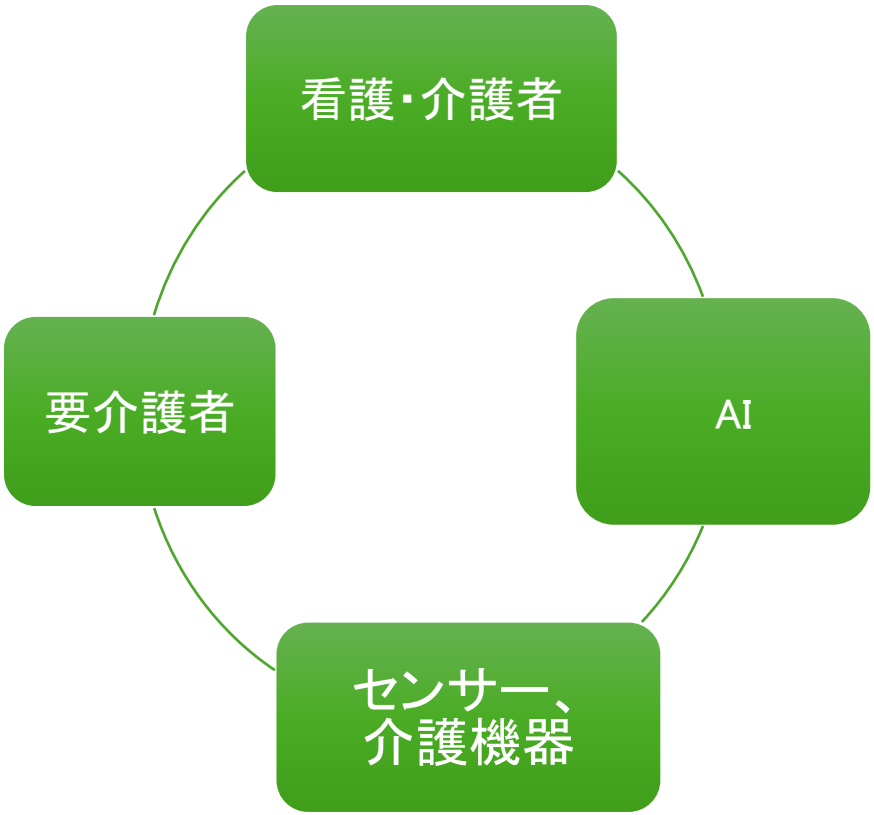
労働からの解放

・介護者の身体的・精神的負担の軽減
・見守り支援（転倒防止）、服薬支援
・要介護者の心理的負担（申し訳ない、恥ずかしい）の軽減
・セラピー支援（会話による認知症等の症状改善）
・感染症の拡大防止

相互性、互恵性
自律という価値

デメリットの可能性
・人手不足の解消や人件費の削減
　⇒介護の質の低下
・巡視の削減
　⇒人間と人間の間の相互関係、互恵関係の縮小
・要介護者の自律・尊厳の軽視
　＝監視されている
・外国人介護者の排除、要介護者へのバイアスの危険

人工物を介したネットワーク



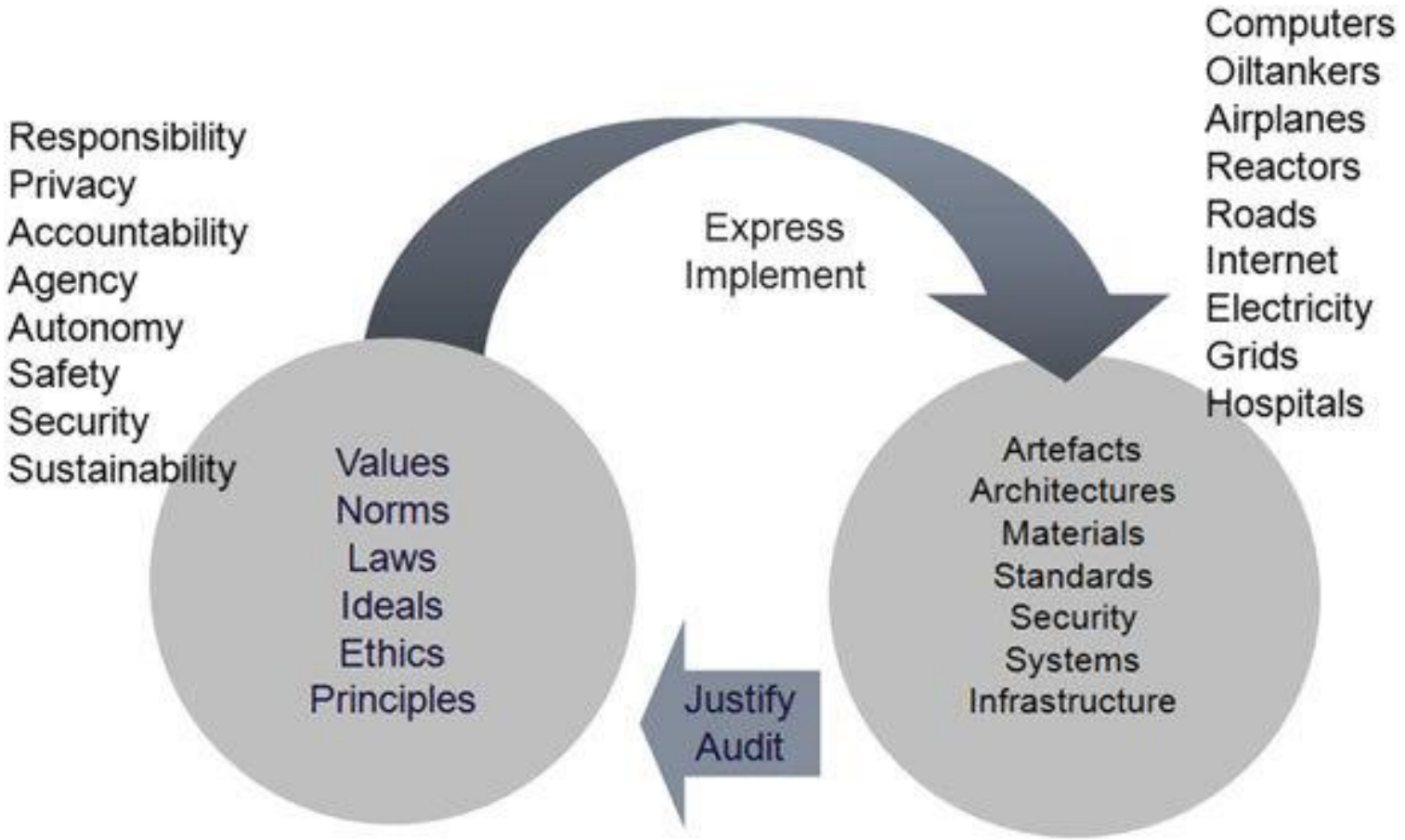
複数の観測者、入力装置、
データ処理コンピューター、
ターゲットとなる患者の器官、
これらを要素とするネット
ワークによって知が営まれる

⇩

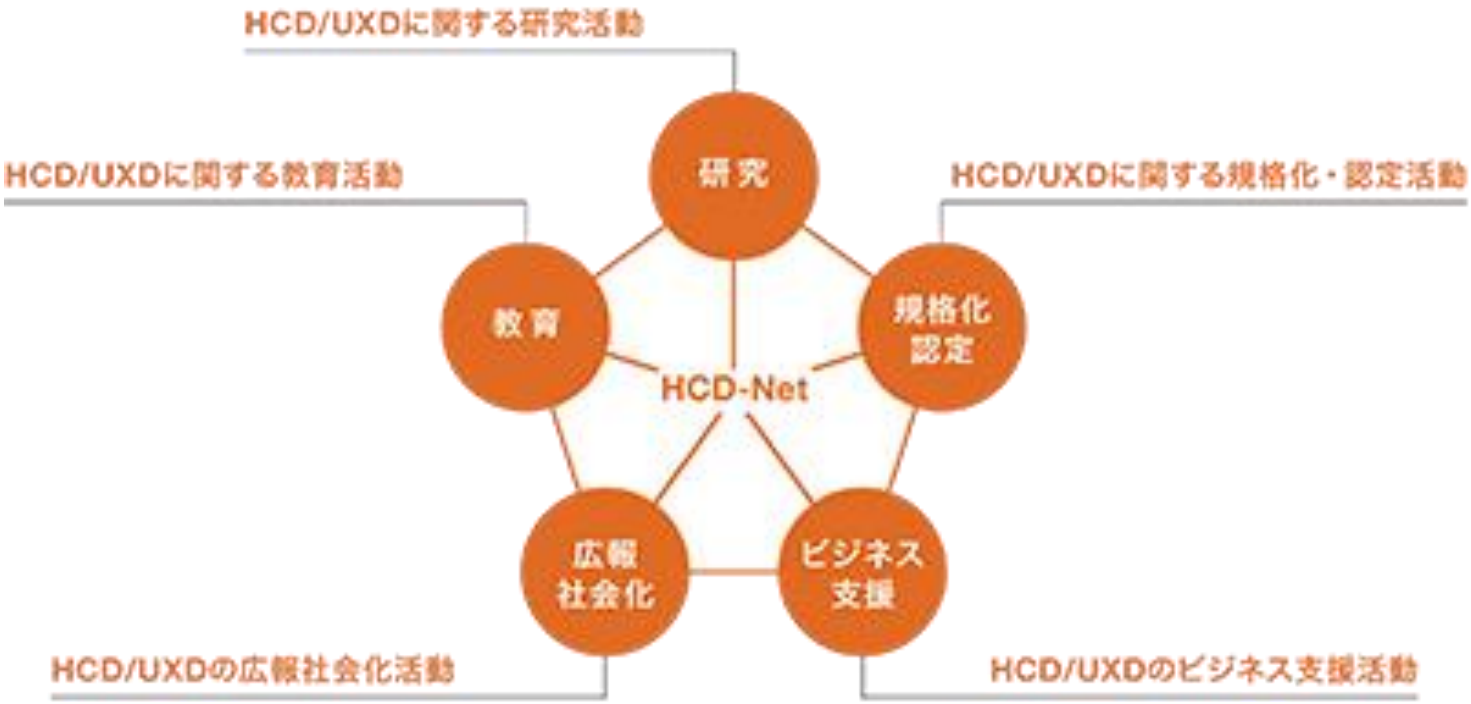
人や制度と一体になった設計

これのみにフォーカス
するとTechnological
Fix

人工物（AIなど）への価値の組み込み



https://learning.edx.org/course/course-v1:DelftX+RI101x+1T2016/homeより



人間中心設計推進機構　HCD-Netのめざす社会

https://www.hcdnet.org/organization/

Technological Fix:気候工学の例

気候工学

気候工学とは、気候変動を緩和するため、気候や大気や海などの地球システムを大規模かつ人為的に操作すること。その代表的な手法は、下表の通り。

分類	**温室効果ガス（CO_2等）の除去** Greenhouse Gas Removal (GGR), Carbon Dioxide Removal (CDR)	**太陽放射管理** Solar Radiation Management (SRM)
概要	化石燃料の燃焼などによって排出された温室効果ガス（CO_2等）を大気中から除去する。効果を発揮するには少なくとも数百年貯蔵されなければならない。	地球に到達する太陽光を減らす、地球の反射率を増やす等の方法によって、地球に出入りするエネルギーの流れを管理することで、熱量を下げる。海洋酸性化対策にはならない。
技術の例	• **海洋肥沃化**（OF: Ocean Fertilization）…海に鉄を散布することなどによって海中生物の光合成を促し、CO_2吸収を進める。海洋生態系への影響が懸念される。 • **炭素回収貯留**（CCS: Carbon Capture and Storage）…化石燃料燃焼時に排出されるCO_2を工学的に回収し、貯留する。回収したCO_2を利用する**炭素回収利用貯留**（CCUS: Carbon Capture Use and Storage）という考え方も。多くのエネルギーとコストがかかる。海洋貯留の場合は海洋酸性化のリスクもある。 • **バイオCCS**（BECCS: Bioenergy with Carbon Capture & Storage）…大気中のCO_2を吸収する植物をエネルギー源として、その際に排出されるCO_2を回収・貯留する（マイナス排出になる）。	• **成層圏エアロゾル注入**（SAI: Stratospheric Aerosol Injection）…成層圏にエアロゾル（硫酸の微粒子等）を散布することで寒冷化の効果を得る。他の気候工学の手法と比べると比較的費用が少なくて済むとの見解もあるが、継続的な実施が必要。SAIに温室効果ガスを減らす効果はないため、いったん開始したあとに何らかの理由で中断せざるをえなくなった場合、短期間で急激な気温上昇を招く恐れがある。 • **海洋上の雲の白色化**（MCB: Marine Cloud Brightening）…雲量を人為的に操作してより多くの太陽光を宇宙に反射し返すようにする。海水を海洋上の雲に向けて吹き上げる等の方法が提案されている。

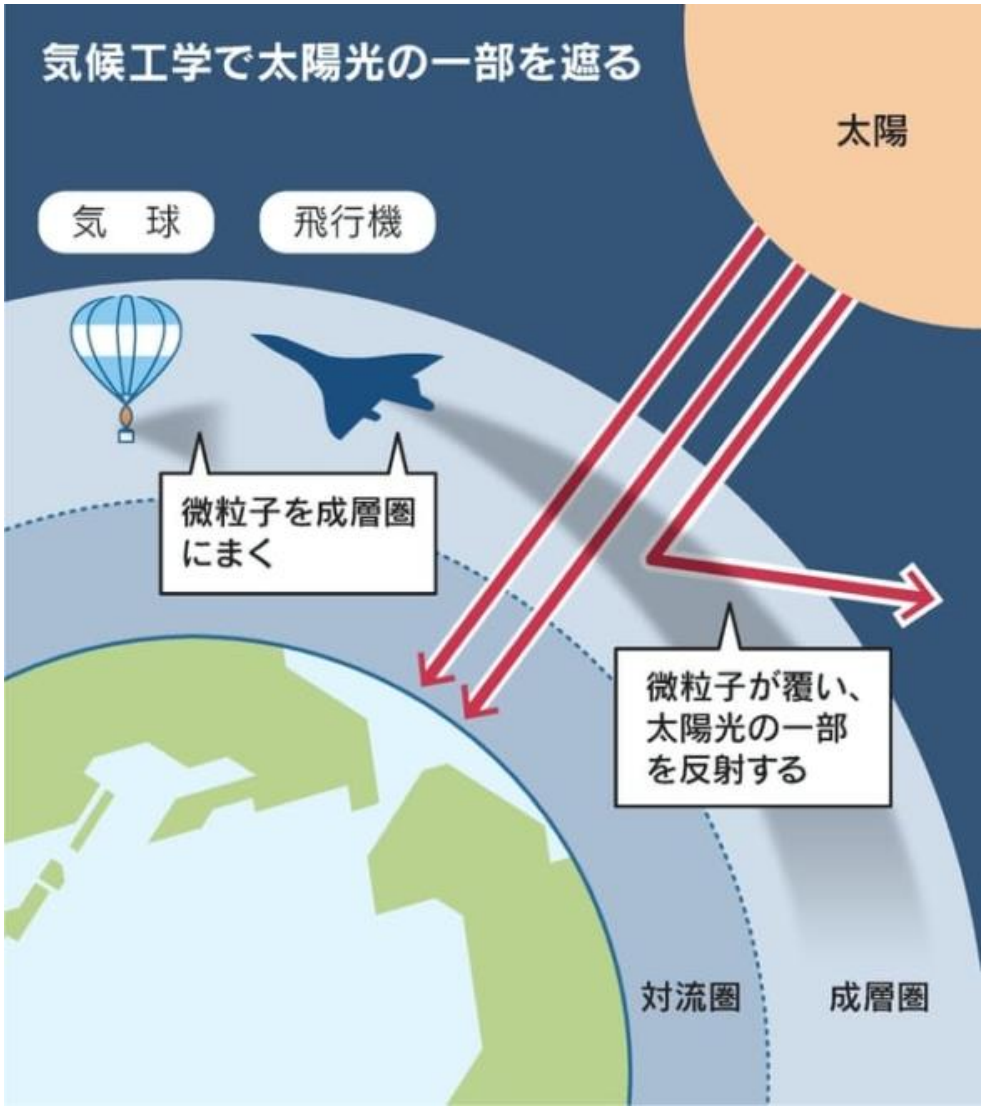
出典：Geoengineering Monitor等より https://kikonet.org/kiko-blog/3469/

成層圏エアロゾル注入(SAI)

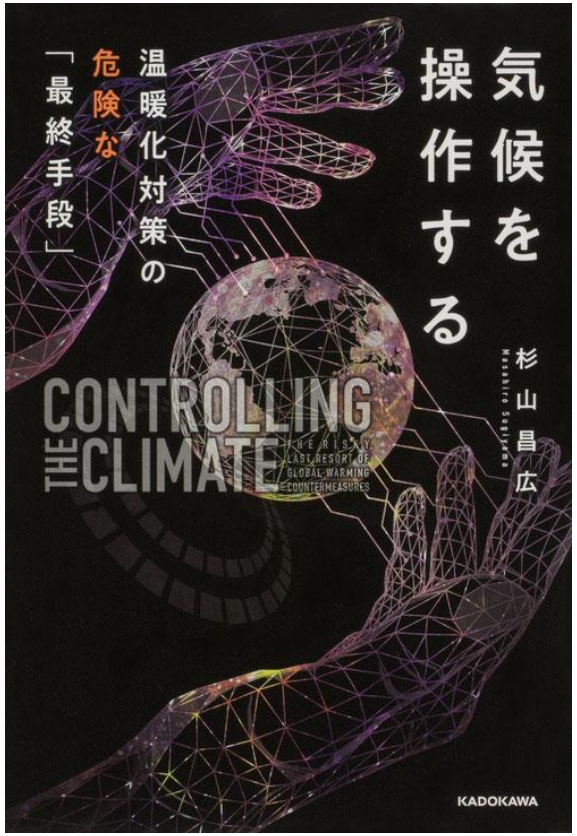
飛行機などを使って成層圏にエアロゾル（硫酸塩などの微粒子）を散布し、雲をつくって太陽光を遮断するものだ。エアロゾルで地表の気温が下がる効果は、火山の噴火で実証されている。1991年のピナツボ山の噴火では、地球の平均気温が約0.5℃下がった。

……理論的には、SAIで20年以内に工業化以前の水準まで地球の平均気温を下げることができる。これによってできる雲は上空約20kmの成層圏に滞留するので、地上に大気汚染は出ない。

（池田 信夫「氷山の溶けるテールリスクに備える「気候工学」」https://www.gepr.org/contents/20210814-01/ ）



日本経済新聞より



太陽放射改変は、技術的にはサイエンス・フィクションではないのです。とりわけ、成層圏エアロゾル注入は自然現象を真似たものです。数十年おきに起きる大規模火山噴火に伴う反射性物質の大気上空への注入、それとともに起こる気候の冷却という一連の観測事実から、科学的に一定の信頼が置けるのです。スーパーコンピューターを用いた気候モデル研究でも同様な結果が得られます。つまり、地球を冷やす事は可能なのです。同時に、冷やし方が均一でなかったり、冷やしすぎた後で突然やめると急激な気温上昇が起こったりするなど、副作用やリスクも明らかになってきています。

そもそも、地球を冷やすことが可能だとしても、それは社会にとって良いことなのでしょうか。地球温暖化に限らず、科学技術イノベーションが加速する中、交流サイト（SNS）は民主国家の選挙に影響を与え始めており、ゲノム編集はデザイナー・ベイビーにつながる懸念もあります。太陽放射改変も同様です。科学的な効果や副作用は見通せても、社会的な影響はまだまだわからないことがあります。

批判や懸念

１　そもそも実現不可能という指摘。現時点では、気候工学は、実施できる段階にないし、そうなる見込みもたっていない。

２　地球・周辺の環境への悪影響。気候工学によって、生態系や人類社会に悪影響を及ぼす可能性がある。特に、SAIは、将来的に何らかの理由で実施できなくなった場合、10〜20年という短期間に急激で破滅的な気温上昇を招くリスクがある（終端問題と呼ばれる）。

３　民主的なガバナンスが困難。ごく一部の先進国のごく一部の科学者や企業の手に地球の気候全体をコントロールしうる気候工学の実施を委ねてよいのか、仮に委ねるとしてもどうやるのか。

４　費用の問題。省エネや再エネは投資回収が見込まれるが、気候工学は持ち出しであって投資回収は不可能。安価な手法もあるとの見解もあるが、比較対象や割引率次第で、一概に安価とは言えないとの指摘もある。また、外部費用（気候工学の実施によって発生する環境への悪影響及びそれに起因する被害への対処の費用等）を考慮に入れれば、費用が想定よりも膨れ上がる恐れもある。

５　国際平和への脅威。天候パターンを改変できるようなある種の気候工学は兵器に転用されるリスクがある。この観点からは、研究を進めること自体にも問題があることになる。

環境倫理とAIによる意思決定の問題

仮想ストーリー 「緑のブレイブ・ニュー・ワールド」

20年ぶりに、再び街を見ることができた。暖かくも穏やかな春の日差しと微風が顔に心地よく、鳥の鳴き声も聞こえてきた。しかし、ほんの数羽ではない。たくさんだ。蝶々もいた。美しい色をしたものが多い。動物も、何種類いるかわからないほど。そして、どこもかしこも緑でいっぱいだ。木々や草木が生い茂り、都市とは思えないほどだった。1960年代からグリーン・ムーブメントが訴えてきたような街でもない。この春は、静かではなく、騒がしく、活気に満ちていた。彼はすぐに、これがもはや彼の知っている世界ではないことに気がつきました。彼が反抗した文明でもない。彼が変えたいと思っていた社会でもない。20年前、彼とエコテロリスト集団がオイルタンカーを爆破して目指した新しい、より良い世界のように見えてきた。結局、反乱軍が勝利したように思えた。気候変動は抑制された。種の大量絶滅は防がれた。生命は勝利した。地球の生態系は回復し、改善さえされた。
やがて彼は、支払わなければならない代償を知ることになる。実刑判決だけでなく彼の若い人生の20年だけでなく反乱軍の仲間の死だけでもない。最大の代償は、人間の自由を失うということだ。皮下に埋め込まれた電子デバイスは、単に彼を追跡し、刑務所での行動をコントロールするためだけではないことがすぐに判明した。この装置は全員に義務づけられていた。

ほとんどの人間は愚かで怠惰であるため、環境破壊、気候災害、大量絶滅を止める唯一の解決策は、人間の行動を直接制御し、操作し、調整することであると主張されていた。人々は、自分たちのライフスタイルを変えることができず、自分たちの生存を脅かしているので、正しいことをするように強制される必要がある、と主張されていたのである。人々は自分自身から救われる必要があるのだ。そして、人間のリーダーシップは非常に信頼性が低く有害であることが証明されており（2020年のコロナウィルスの大流行のような危機の時にあまりにも明らかになった）、膨大な量のデータとシステムの複雑さを考えると複雑な気候問題には全く対処できないため、人工知能（AI）に地球のコントロールを委ね、人間の代わりに意思決定を行うことに決定されたのである。政策決定だけでなく、多くの個人の意思決定も機械に委ねられるようになった。水の使用量、食事の選択、交通手段の利用を規制していた柔らかな女性の声は、刑務所の外でも彼に同行していた。

２つのストーリー

人々、企業、国家、その他のアクターの役割は？

必須の要求：地球温暖化の回避

自由放任主義
人々、企業、国家、その他のアクターが気候変動に対して何をする（しない）かについて自由にしておくべき

デジタル権威主義
生命、財産の保持のため、権威主義的な国家が、さまざまなアクターがどのように行動すべきかを統制する

（グリーン・リバイアサン）

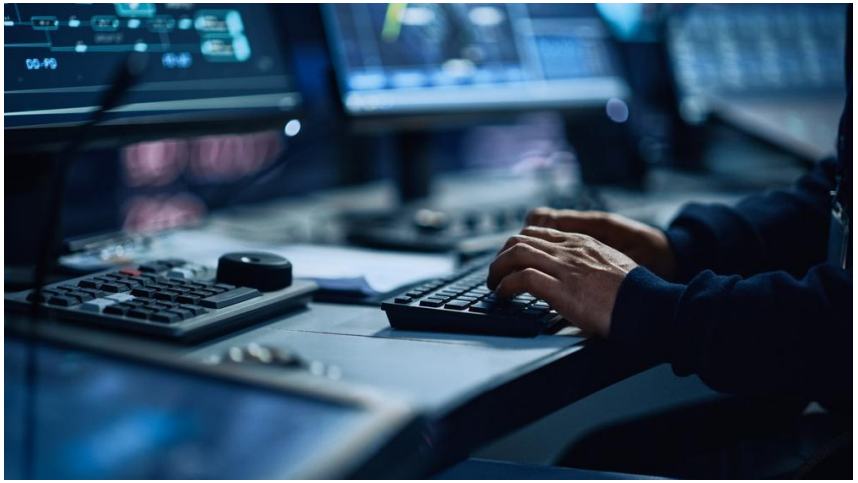
自由と民主主義の放棄を受け入れるべきか否か

「デジタル権威主義」
政治的課題の一つは自由に関するものである。もし私たちの苦境が本当にひどいものであるなら、私たちは人々、企業、国家、その他のアクターが気候変動に対して何をする（しない）かについて自由にしておくべきだろうか、それとも行動を強制すべきだろうか、するとしたらどのような方法でだろうか。権威主義的な国家が、どのように行動すべきかのモデルとなるべきなのか、国家が統制した行動が気候危機に対処するのにより効果的であることを考えれば、そうすべきなのか。それとも、自由主義的な政治をモデルとし、行動を個々の行為者に大きく委ねることで、高度な自由を維持しつつ、変化が起きるかどうか、あるいは悪い方向にしか変化しないかどうかを保証するべきだろうか。

もし、企業が「地球のために」AIを開発し、持続可能で気候に優しい方法でAIを利用するという主張が、「倫理的洗浄」、つまり従来通りのビジネスを行うためのイチジクの葉（形だけ飾り）に過ぎないとしたらどうだろう。もし、企業が変化しても、その効果が非常に小さいとしたらどうだろう？気候（とAI）のグローバルガバナンスのための効果的な政治制度が欠けているため、ほとんど起こらないとしたら、つまり、自由度が高すぎるとしたらどうだろう？また、AIによる気候変動の緩和策が、権威主義的で自由を脅かす傾向と結びついた場合はどうだろうか。
(Mark Coeckelbergh, Green Leviathan or the Poetics of Political Liberty,2021)

類似課題　AIと自由

例）コンテンツ・モデレーション（投稿監視）

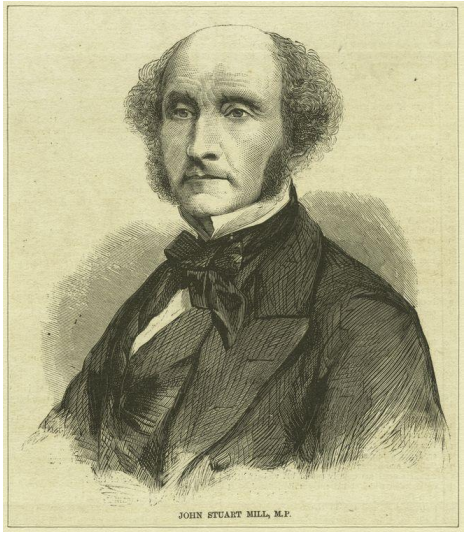


・ソーシャルネットワーク上には膨大な量の偽情報や、性差別や人種差別、暴力や悪意を含んだコンテンツ、児童ポルノや過激な暴力画像などが存在
・不適切なコンテンツの拡散を防止し、プラットフォームの信頼性を確保するために、AIを用いて問題となりうるコンテンツを検出し、そのコンテンツを削除ないしラベリングするコンテンツ・モデレーションを実施。
・人間のモデレーターと違い過度な精神的ストレスを与えるおそれがない。

表現の自由の価値
・議論を通じて真理を見いだすためには、言論の自由が必要だ。人は誰もが誤りを犯す可能性を持っている。思想の自由なやり取りは、真理が現れ、独断的な信念が疑われる機会を増やす

ジレンマ
1、ヘイトスピーチや操作、偽情報は、政治参加や政治的自己実現の自由、人の道徳的尊厳、そして共通善を危うくする
2，言論の自由の侵食を許してしまうと、権威主義的な政権が民主主義を破壊することを容易にする

古典的な自由主義哲学（ミル）では、このジレンマは、言論の自由および干渉からの自由そのものに関するジレンマとなる



共時的倫理（環境正義）と継時的倫理（世代間倫理）

	世代内倫理 （共時的倫理）	世代間倫理 （継時的倫理）
格差の例	一人当たり資源消費の南北格差	自然資源を享受する機会の世代間格差
受苦の例	・マイノリティ居住地域への有害廃棄物処分場の立地 ・原発の「過疎地」への立地 ・公害や核被害の子どもや（被曝）労働者へのしわ寄せ ・公害輸出 ・開発搾取	・核廃棄物管理の子孫への強制 ・地球温暖化

Webinars Around the Earth
WHY DO WE NEED
GREEN RIGHTS ?
Launch of the Green Rights Coalition
Prof. Nicholas Robinson
Prof. Victor Tafur
Tony Oposa
Victoria Lichet
I C E L
International Council of Environmental Law
GreenRights Coalition

「リスク配分の正義」

・リスク計算は「配分の正義」を犯す。

コストベネフィット分析や「最大多数の最大幸福」（ベンサム）では、弱者がリスクを引き受けることになる。

・平均的閾値では、受容可能でない人々が出てくる可能性。

あらゆる人々を同等に保護するのではない

参考　ベック『リスク社会』：後期の産業社会では、産業と科学による環境や健康のリ スクの増大に伴って、階級を基盤にした従来の富の分配をめぐる争いに 代わり、リスクの分配をめぐる争いが主要な社会的対立となった。

「ブーメラン効果」：原発の重大事故や生態系破壊のリスクのように、リスクは誰かの人間の行為がつくり出したものである。リスクの原因者とリスクを被る者とは一致しない。

多様な価値の間のバランス

プライバシーのための設計
セキュリティのための設計
包含のための設計
持続可能性のための設計
民主主義のための設計
財産や安全性のための設計
透明性のための設計
説明責任のための設計
ケイパビリティのための設計

・人権や持続可能性、公正さ、安全性は一つのデザインにおいて適合されうるのか？
・人工物に組み込まれる価値の重みをだれが決めるのか？

人間中心の設計
多様なステークホルダーの参加

古くて新しい責任分散の問題

2010年のメキシコ湾原油流出事故

油まみれのペリカンたち

海岸に流れ着いた原油	海面に広がる原油

・2010年4月20日、アメリカルイジアナ州ベニス沖にある原油掘削施設「ディープウォーター・ホライズン」において、大規模な暴噴事故が発生し、噴出した油が爆発。この爆発事故により、そこにいた126名のうち11名が行方不明。掘削施設は二日間燃え続け、22日に沈没。 しかし、掘削施設沈没後も海底にはパイプが残り、その破損箇所などから原油の流出が続き、それがメキシコ湾一面に広がる。
・映画「バーニング・オーシャン」(2017)

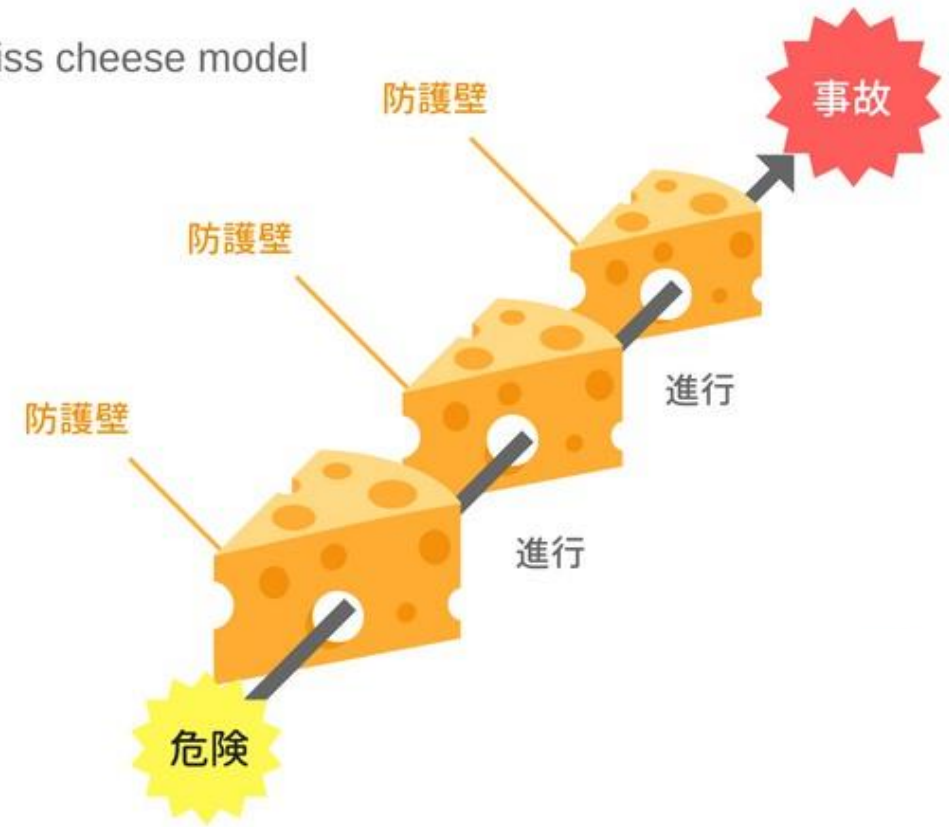
政府はBP社に浄化の管理責任と被害者への補償責任を負わせたが、さまざまな調査や法的手続きによって、関係する複数の企業や個人のうち誰が流出事故とその影響に責任があるのかが問われている。

Many hands 問題

ある出来事に対して集団（ネットワーク）としては道徳的責任があるが、その集団（ネットワーク）を構成するどの個人ないし団体にも道徳的責任がないようにみえる事態。

複雑なエンジニアリングプロジェクトでははは、責任の所在を明らかにすることが非常に困難な場合が存在する。（状況の複雑さと関係者の多さのために、誰かに合理的な責任を負わせることが困難ないし不可能である。）

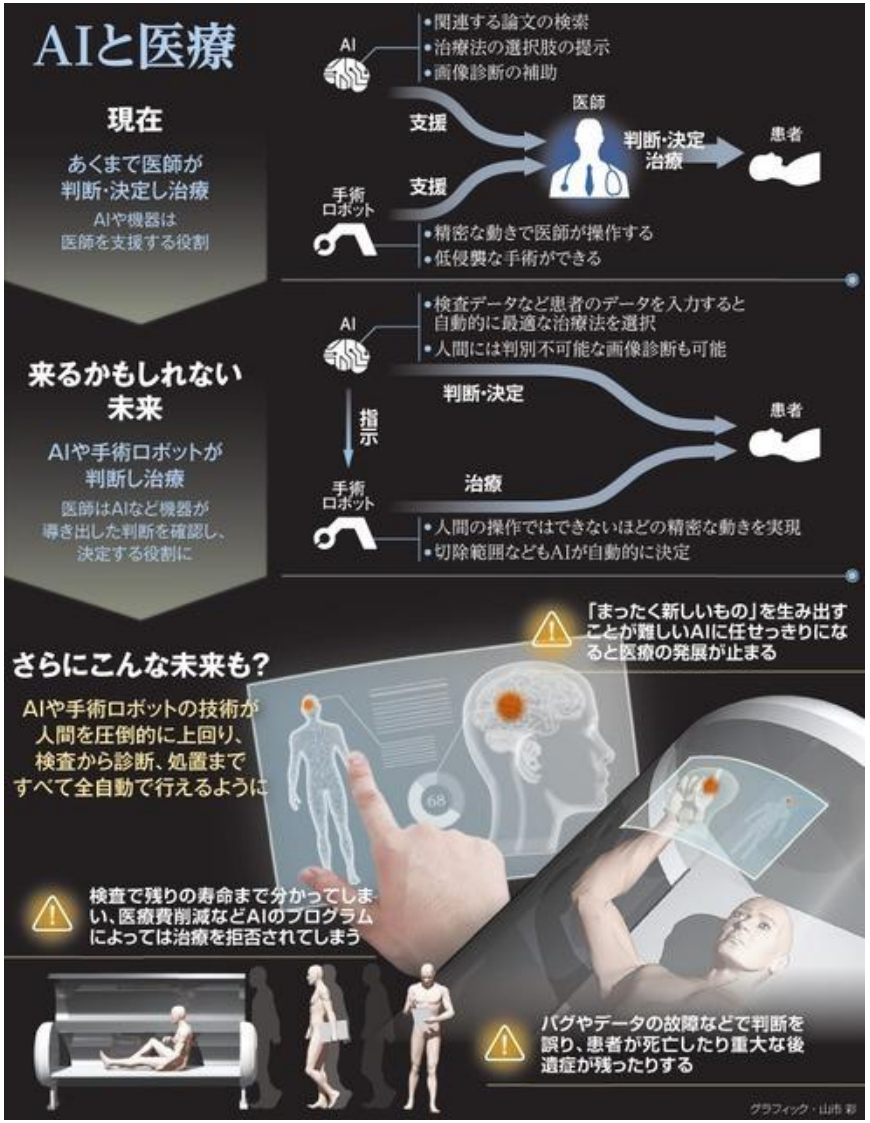
スイスチーズモデル
事故は単独で発生するのではなく複数の事象が連鎖して発生するとしたもの



もしも、AI医療で手術ミスが起きたら？

2017年　厚生労働省「保健医療分野におけるAI活用推進懇談会報告書」
「診断確定や治療方針の最終的な意思決定は医師が行い、その意思決定の責任も当該医師が負うべき」
（確定診断そのものをAIが行うことを否定。確定診断は医師の役割）

2019年　厚生労働省「保健医療分野AI開発加速コンソーシアム 議論の整理と今後の方向性」
「AI技術の進歩は目覚ましいことから、人間とAIの協働などに関する他の分野での議論や技術的進展を見ながら、今後も（このような建前を継続するべきかについて）継続的な議論が必要」



自動運転はだれに道徳的責任が求められるのか？

人間：意識と自由を持ち、道徳的・非道徳的な行動、推論、判断、および意思決定を行う。道徳的行為者

自動運転車：道徳的な行動、推論、判断、および意思決定を組み込むか？
道徳的行為者？

従来型の自動車。人間の殺傷という非道徳的な結果をもたらしうる。道徳的行為者ではない。運転者が道徳的行為者。

（https://toyokeizai.net/articles/-/228800）

自律学習型のインテリジェントシステムでは誰が責任を取る？

もし自ら目標を設定できるロボットが殺戮の限りをしたら、私たちは誰を非難するだろうか？あるいは、そのようなロボットが、私たちの理解する道徳的に許される行動と矛盾するやり方で自律性を行使したとしたら、私たちは誰に正当な責任を負わせることができるだろうか？このような自律性を持つロボットを作る可能性は、私たちにもあるのだろうか？……このような軍事兵器が、直接的または間接的な指示なしに、何らかの形で独自の決定を下すことができるというシナリオを想像することができる。そのロボットは、他者から指示された目的を追求するために行動するかもしれないが、予測不可能な（予測不可能な）方法で行動し、道徳的に許容されるその目的を追求することには沿わない。

(Marc Champagne & Ryan Tonkens, 2013)

・地球温暖化は喫緊の課題

・他方、人間とAIの協働関係のなかで、デジタル権威主義にならない意思決定のあり方が必要。

⇒民主主義的で人間中心的な協働関係

技術の公共的決定

- 技術は設計者の意図を超えた結果（作用・副作用を）をもたらす
- 人工物にどのような価値を持ち込むと人々に善をもたらすかは、技術者の側でパターナリズム的に決定することがむずかしい
- 技術的観点にかぎらないより多くの観点から、どのような人工物を作るかは開かれた場で決められるべき（事物をめぐる政治）

価値発見的な視点とは（再掲）

- 新たな社会を設計し、その社会で新たな**価値創造**を進めていくためには、多様な「知」が必要である。特にSociety 5.0への移行において、新たな技術を社会で活用するにあたり生じる**ＥＬＳＩ**に対応するためには、俯瞰的な視野で物事を捉える必要があり、自然科学のみならず、人文・社会科学も含めた「総合知」を活用できる仕組みの構築が求められている

『第6期科学技術・イノベーション基本計画』（2021年）、14頁