公益財団法人
未来工学研究所
INSTITUTE FOR FUTURE ENGINEERING

知識論と学問論を踏まえた新型「ソブリンAI」：知能型から知性型への進化

2026年8月7日

未来工学研究所　平澤　泠

過去のレビューシリーズの資料と動画サイト
https://www.ifeng.or.jp/2024_grips_lecture/

Copyright © INSTITUTE FOR FUTURE ENGINEERING

Sovereign AIに関するレビューシリーズ（１）

■主題

- 「ソブリンAI」に関する主要国・地域の取り組み状況の把握
- 「ソブリンAI」に期待される機能
- 「ソブリンAI」のあるべき構成と構造
- 「ソブリンAI」に関する上記以外の事項（今後の検討の中で提起される上記以外の緊急事項）

■Sovereignの語源から位置付ける本シリーズの焦点

- Sovereign ← 俗ラテン語superanus「最高位にあるもの」
- 本シリーズでは「**最上位の政策や戦略に関わるAI**」と理解する

Copyright © INSTITUTE FOR FUTURE ENGINEERING

未来研におけるAI調査研究

公益財団法人
未来工学
研究所
INSTITUTE FOR FUTURE ENGINEERING

■知識論・システム論をベースにしたAI関連調査

- 「次世代AIとは：認識論・知識論・ソフトサイエンスからのアプローチ」コアタイム講演 2017.09.06

■新技術振興渡辺記念会

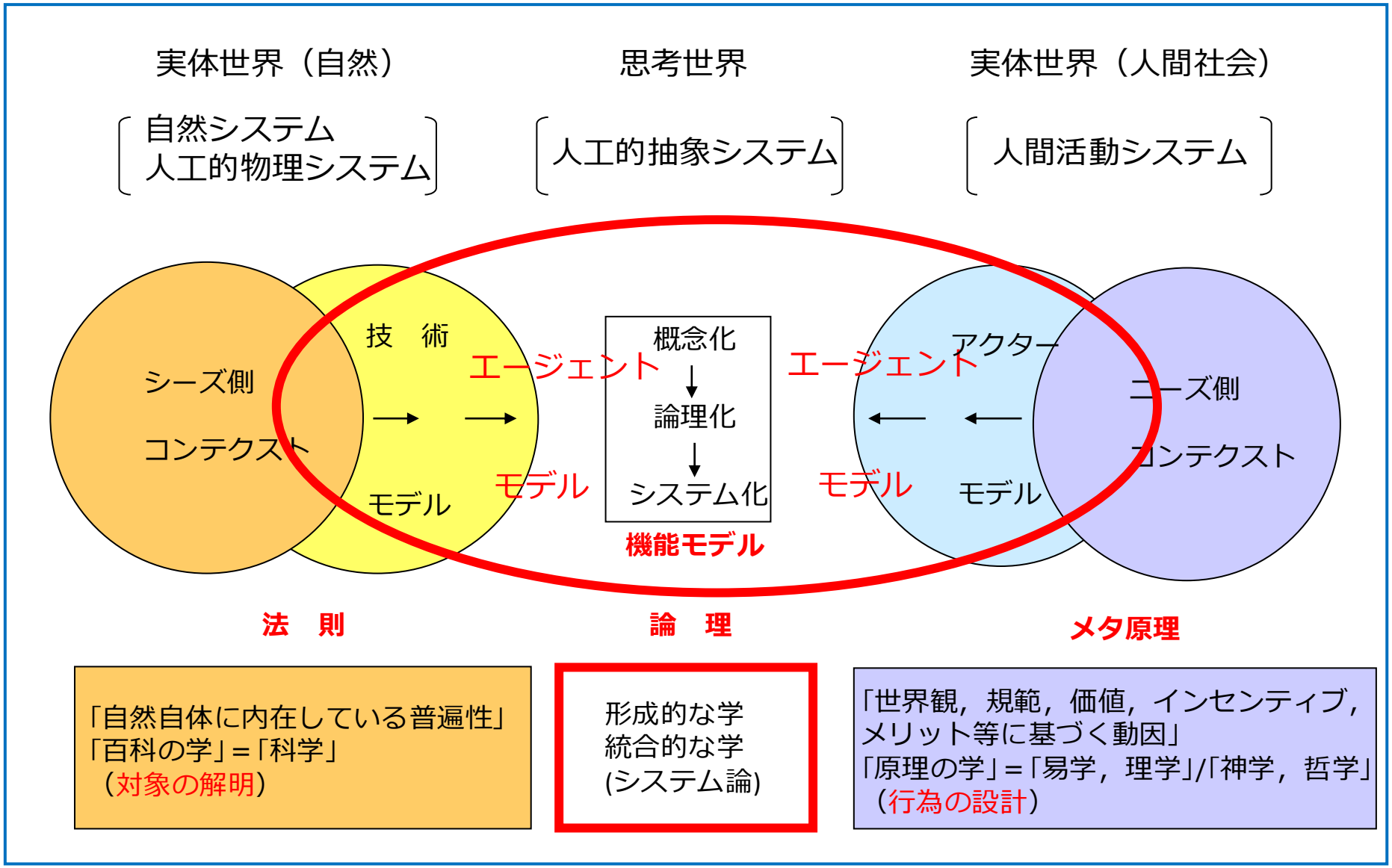
- 「ディープラーニング（深層学習）を中心とする第３世代型人工知能の限界を探る調査研究」2018.5（未来工学研究所）
- 「次世代人工知能の構想Ⅰ」2019.9（未来工学研究所）

■特許申請活動

- パーソナルAI　2023.07.31

Copyright © INSTITUTE FOR FUTURE ENGINEERING

知識の概念モデル（科学主義）

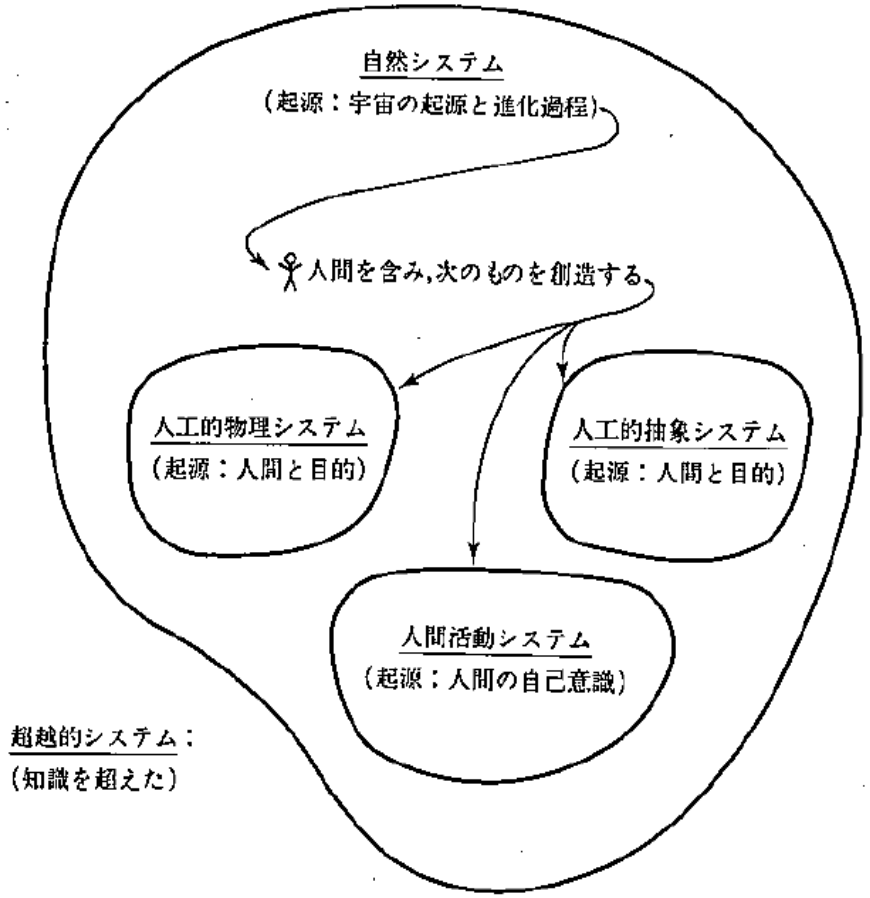


Copyright © INSTITUTE FOR FUTURE ENGINEERING

認識の類型

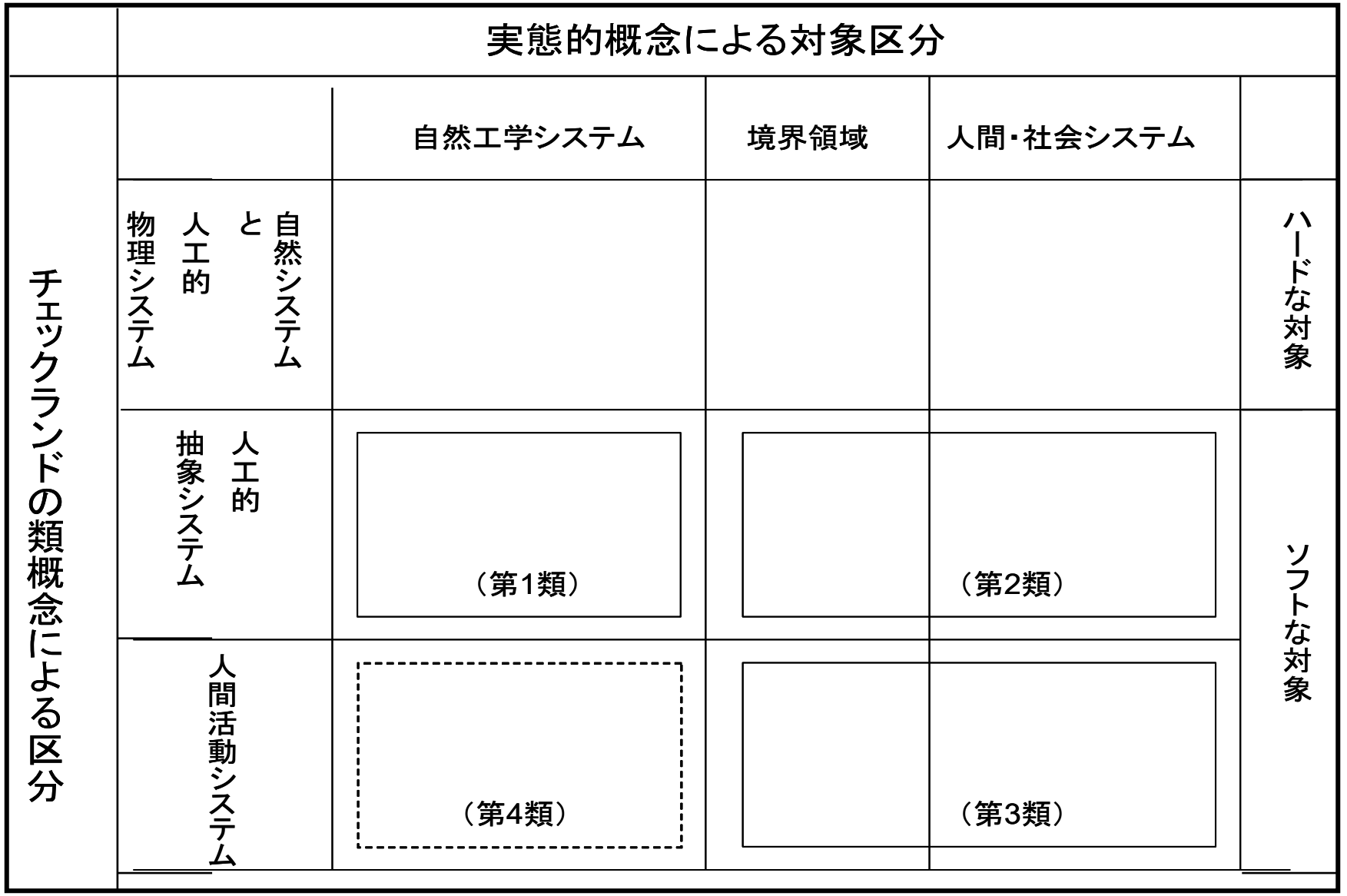
◆**自然システム**： 自然の摂理のままに支配されている実体的対象や現象
　あるがままの自然の中には、生理学や行動学の対象となる人間も含まれる
◆**人工的物理システム**： 人間によって作られた実体や現象
　機械や工場等
◆**人工的抽象システム**: 人間の内的過程を経て構想された対象や現象
　思考世界の産物
◆**人間活動システム**： 人間の自己意識に基づく活動
　意図的行動等
◆**（超越的システム）**： 人間の実体的知識を超えたシステム
　神、あの世等
　＊）ピーター・チェックランド、「新しいシステムアプローチ」、オーム社

■**ソフトな対象**
- **第１類**： 自然システム・人工的物理システムを対象
  気象予報ソフト、装置制御ソフト等
- **第２類**： 意思的人間を含む対象
  GDP成長率予測ソフト等
- **第３類**： 人間活動システムを対象
  ブレーンストーミング手法等
- **（第４類）**： 人工生命？



Copyright © INSTITUTE FOR FUTURE ENGINEERING

ソフトな対象



＊）　「ソフト系科学技術の研究開発の現状及び今後の展開方向について」、政策科学研究所（1988）

Copyright © INSTITUTE FOR FUTURE ENGINEERING

一般システム論

■ルードイッヒ フォン ベルタランフィ（長野敬、太田邦昌訳）「一般システム理論」みすず書房（1973）
存在するもの全体を考察の対象に置き、そこに通底する理論を見出すことを願って始められた。ISSS
自然科学の対象領域では対象固有の分析的方法論が効果的に知見を生み出したのに対して、システム論では対象を構成する**要素の関係性**に着目して、（要素の内部に分け入るのではなく）全体を包括的に把握することにより新たな知見を生み出すという方法論に独自性がある。

宇宙の起源と生物の進化過程を経てヒトが登場するまで世界は「自然システム」のみで構成されていた。ヒトは自己意識を抱き個別の価値観や情動に促され自身の存在に基づき「人間活動システム」を構成し、また「自然システム」 に手を加え「人工的物理システム」を創出した。そしてこれら3システムに対する思考世界「人工的抽象システム」を想起する。さらにこれら知識に根ざし たシステムには依拠しない「超越的システム」を創造することもあった。

	意志的人間不存在	意志的人間が存在
実体世界	自然システム（ハード1） 人工物的物理システム（ハード1）	人間活動システム（ハード2）
思考世界 （人工的抽象システム）	人工的抽象システム1（ソフト1）	人工的抽象システム2（ソフト2）
		人工的抽象システム3（ソフト3）

Copyright © INSTITUTE FOR FUTURE ENGINEERING

科学哲学/認識論/知識論

科学哲学を基底とする認識論の枠組みによる思考過程は、対象を**外部から観察**し、その対象に内在する普遍的原理、つまり法則を見出そうする。この枠組みの下で得られる知識体系は、自然によって保証された**「確かな根拠」**によって支えられている。観察者の願望や価値観とは無縁のものである。観察者の「確信」は「確かな根拠」（アンカー）には当たらないと考える。
戸田山和久「知識の哲学」産業図書（2002）

さて、宇宙を構成する知のカテゴリーに関し科学哲学を基底とする認識論の枠組みで分類し（要素に分け）それぞれのシステムの関係を明らかにした。

- **自然システム**：普遍的（どこをとっても何時においても）に内在原理（法則性）がある
- **人工的物理システム**：ヒトによって構築されたが、普遍的内在原理が存在している。自然システムと人工的物理システムはこの枠組みの中では同一の挙動をとるので、合わせて「**ハード1**」と区分する。
- **人間活動システム**：価値観や情動に促され個別に独自の行動を行っているヒトを含むシステム。これを「**ハード2**」と区分する。
- **人工的抽象システム**：ヒトは「思考世界」を構想する。「実体世界」はヒトの認識の外に存在しそれを「**外界**」と言うとすれば、思考世界はヒトの認識の内「**内界**」に存在し、「モデル」としてそれを外部に表示することが出来る。自然システムや人工的物理システムを対象にして思考世界で構想したモデルは「**ソフト1**」と分類する。「ソフト1」は実体世界に照らし合わせモデルとしての確からしさを確認できる。仮想的な価値観や同 一の情動を有するヒトの集団を想定し思考世界にモデルを描いた場合それは「**ソフト2**」と分類する。「ソフト2」の確からしさは実体世界に照 らして確認することは困難である（そのような実体は通常存在しない）。さらに、独自の価値観や情動に導かれて行動するヒト（ないしヒトの集団）を思考世界で描いた「モデル」は「**ソフト3**」と分類する。「ソフト3」の確からしさを確認することも困難ではあるか不可能ではない（現に存在しているが、個々に異なり（不均）変動し移ろいゆく（不定）ところに困難さがある）

ヒト自体は幾つかのシステムの構成要素になる。医学や生理学の対象となる人体（自然システム）、標準的な価値観を有する組織や社会の構成要素であると仮想する経済学モデル（人工的抽象システム：ソフト2）、価値観や情動に導かれて躍動するヒト（ハード2、ソフト3）

Copyright © INSTITUTE FOR FUTURE ENGINEERING

報告書1：「ディープラーニング（深層学習）を中心とする第３世代型人工知能の限界を探る調査研究」

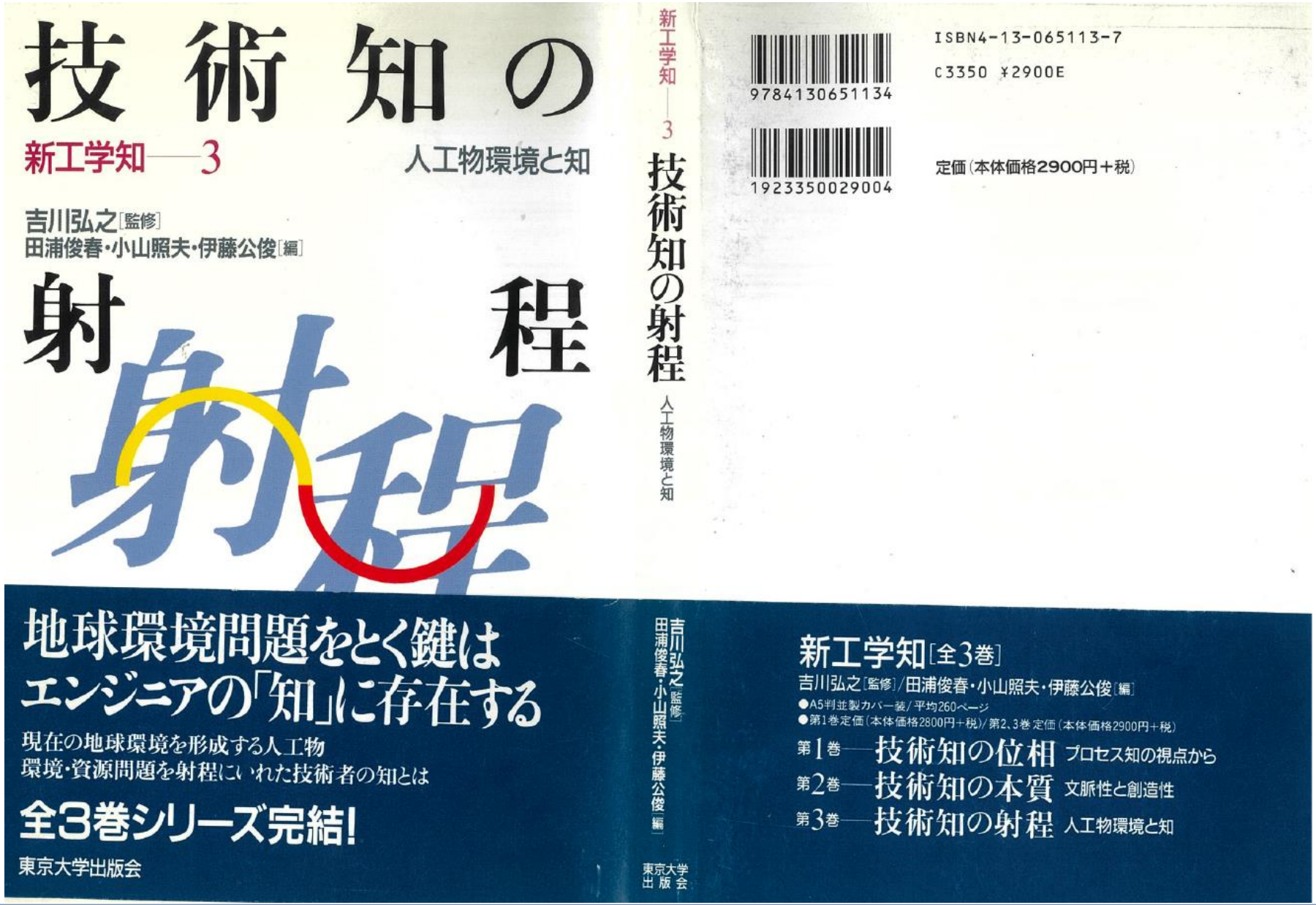
知識カテゴリー	構成システム	事例的説明	特徴
ハード1	自然システム、人工的物理システム	自然や機械	対象に普遍的に内在する法則がある
ハード2	人間活動システム	意志や価値観に従う人間を含む	内在する法則はない。対象により異なり時と共に移ろい行く。経験的原理の探索
ソフト1	人工的抽象システム1	自然や機械を外部から模したモデル	シミュレーション結果を照合する実体に法則性がある
ソフト2	人工的抽象システム2	人の行動や組織を疑似的に外部から模したモデル	対象とする実体は不均一で不定であるが、疑似的な法則性をたててもした論理モデルを代用する
ソフト3	人工的抽象システム3	意志や価値観を含む内面を模したモデル	普遍的、確定的に模したモデルとして把握することは困難。経験的原理の探索
その他	超越的システム	上記の知識カテゴリーや構成システムに分類できない超越的概念	

知識カテゴリーに注目して事例群を分析する。

- 2017年日本経済新聞記事からAI等のキーワードを含む804件
- 2017年人工知能学会誌所載論文論説121篇
- Preferred Networks社のホームページに開示され実施事例71件
- 政府の取り組み事例であるNEDO採択テーマ57件

PFNとNEDOの実施事例はいずれも、 「ハード1」ないし「ソフト1」に属する対象を扱っていて、 学会誌ではさらに「ソフト2」まで対象を広げているが「ハード2」や「ソフト３」 （人文学の課題）の領域には明示的に進入を避けている。 これに対して、 報道記事では全ての知識カテゴリ ーを扱っていて、 DLやAIの現状に対して、 原理的に過大な期待を寄せている様子が読み取れる。

Copyright © INSTITUTE FOR FUTURE ENGINEERING



Copyright © INSTITUTE FOR FUTURE ENGINEERING

刊行にあたって ——— iii

第1部 環境問題と知 ——— 1

メタテクニカ 新時代の哲学的発言 ——— 今道友信 3
環境知のダイナミックスとインセンティブ連鎖の設計 ——— 平澤 冷 29
プロセス知の欠陥と環境問題 ——— 田浦俊春 41

第2部 環境問題と情報メディア ——— 59

環境・資源情報の重要性 ——— 小山照夫 61
複雑現象予測の可能性とその実践的意味 ——— 上草貞雄 79

環境知のダイナミックスとインセンティブ連鎖の設計

平澤 冷

近代科学技術に対置すべき知のアプローチ

近代科学の認識論は、意思的存在としての人間を形式知からなる知識体系に外化できたであろうか。P・チェックランドのシステム・カテゴリーの概念である「自然システム」（医学や生理学の対象となる人間はこれに属する）や「人工的物理システム」（機械や工場など）と異なり、意思的存在としての人間（これは「人間活動システム」に区分される）は、人文学（humanities）の対象であり、〝科学〟的認識対象の範疇には含めていない。チェックランドによれば、システムにはじつはもうひとつのカテゴリーがあり、それを「人工的抽象システム」（たとえば経済のマクロモデル）とよんでいる。[1] 人間科学や社会科学の〝科学〟たる所以は、「人工的抽象システム」として人間や社会をとらえるところにある。この差異は、後に述べる。

意思的存在としての個人を〝科学〟的に正当にとらえられない場合、そのような個人を構成要素とする人間組織や人間社会もまた当然、近代科学の認識論の範疇からはずれる。また、そのような個人が関わる社会的な活動もすべて、正当的には近代科学の認識論ではとらえつくせないことになる。環境問題も通常このようなものであ

公益財団法人 未来工学研究所 INSTITUTE FOR FUTURE ENGINEERING

Copyright © INSTITUTE FOR FUTURE ENGINEERING

近代科学の認識論では、認識対象を対象化するところからはじめる。そして、対象を知るために、まずは現象論的な把握を行ない、その現象のもととなる本質の解明へと進む。現象の本質は、トートロジーに陥らないためには、通常対象そのもののレベルではなく、その下位のレベルの知見によって説明される必要があり、そのためには対象を要素に分割し分析を深めることとなる。そして、そのようにして得られた知を確かなものとするためには、認識主体の確信（内的妥当性）にとどまることなく、その確信された内容を認識主体と切りはなし、その内容を対象によって語らせる必要があり、そのような手続きを経てはじめて、科学的知として認識されたこととなる。対象によって語らせることは、観察者によらずだれがやっても同一の結果が得られることを意味し、これは別の言い方をすれば、対象に内在する原理にしたがって操作されたことにほかならない。その原理は、もっとも厳しく言えば、時空の差異を超えた普遍的内在原理であり、どの対象にたいしても、またいつにおいても成立するものである。われわれは、再現的実証性を積み重ねることにより、そのような原理を認識することとなる。

「自然システム」や「人工的物理システム」には、いわゆる自然の摂理に相当する普遍的原理がその対象を貫いて存在していると考えられる。それにたいして、意思的存在としての人間は、個別的であり、またうつろいゆく存在でもある。そこには普遍的内在原理は存在しない。いや、正確には、存在しないと考える立場から新しい視点を導こうとする。もちろん、意思的存在としての人間においてすら普遍的内在原理が存在するとして、それを追い求める立場も許されるが、私にとってそれはむなしい。さらには、仮想的な原理の存在を仮定し、その仮想的原理を実態に近づける努力を重ねる立場もある。「人工的抽象システム」として扱うのは、そのような立場である。

そもそも、「人工的抽象システム」は現実世界そのものではなく、思考世界を表現するためのものである。したがって、設定された「人工的抽象システム」の内部で論理的整合性がとれている場合、それは〝真〟であるとされ、シミュレーションによる再現的〝実証性〟の確保により、そのことが確認される。数学は本質的にこのようなものであり、公理系のなかの無矛盾性を問うが、現実世界との整合性は、第一義的には問わない[2]。「人工的抽象システム」は、思考世界のなかで、自由に設計できるため、ソフトな世界である。それにたいして、「自然システム」や「人工的物理システム」は、自然の摂理にしばられた、ハードな世界を表現している[3]。人工物にも、知の原理の異なる二つの世界があることに注意すべきであろう。また、環境問題は、通常「人工的抽象システム」の形式で表現されていて、意識の有無にかかわらず、これらは仮想的ないし近似的な議論となる。

さて、意思的存在である人間を含む系（「人間活動システム」）にたいしては、いかなる知が想定されるであろうか。前述のように、普遍的内在原理がそこには存在しないとして、近代科学の所産である科学的知に対置すべきあらたな知へのアプローチを、以下に描いてみたい。

近代科学のアプローチを、前述のように、分析的、要素還元的、そして再現的実証性によって特徴づけた。ここでは、その対極を想定してみよう。分析的にたいし形成的、要素還元的にたいし全体性、総合性をあげることができる。われわれは、工学をはじめとする実学の分野で形成的ないし、構築的アプローチをすでにとってきた。また、対象の〝部分〟ではなく対象の〝全体性〟を考察の対象とする伝統的なシステム論を知っている。しかし、ここでは、あらたな知への試みとして、それらではなく、観察者と観察対象の総合、つまり認識主体と認識対象をわけるのではなく、その両者を含む全体を対象とした知を論じたい。

近代科学が、その対象を対象化し、観察主体と切りはなすことによってあらたな実証性を確保し、ゆるぎない知の体系を築いてきたことをすでに述べた。これを外在的アプローチ（exclusive approach）とよぼう。それにたいしてここで試みることは、認識主体を内在化させる内在的アプローチ（inclusive approach）である[4]。近代科

公益財団法人 未来工学研究所 INSTITUTE FOR FUTURE ENGINEERING
Copyright © INSTITUTE FOR FUTURE ENGINEERING

AI開発のフロンティア

■生成AIの発展段階：ANI、AGI、ASI
- Narrow ←
- General
- Super

■ 知能か知性か
- 知識の量と正確さ
- 知識以外の人間力としての**感性**と**品性**

■ 知恵 Wisdom
- 知識（知ってること）の**正確さ（質）の向上**

■ 知識の正確さ
- 根拠の普遍性
- 確信の確からしさ、共感できる、価値の共有
- モデル内での無矛盾性、限定された・前提の下での意味

■ 誤謬の排除
- 知識論と知識を体系化した学問論を踏まえる

Copyright © INSTITUTE FOR FUTURE ENGINEERING

CATEGORIES OF KNOWLEDGE

３種の知識の確かさ

認識論を踏まえた知識論にもとづく知識の確かさの弁別。根拠の質の違い、学問の成立基盤の違いに着目。

科学的・普遍的知識

自然科学・工学で得られる普遍的知識

検証可能な根拠を有する

内省的・固有的知識

哲学・芸術論で展開される主観的知識

当事者の確信は得られる

集合的・社会的知識

組織論・経済学で論じられるモデル依存的知識

モデル内における無矛盾性

Copyright © INSTITUTE FOR FUTURE ENGINEERING

Copyright © INSTITUTE FOR FUTURE ENGINEERING