

構築的技術哲学としての
クリティカル・ロボティクス
—ELSI・RRIを超えて—

呉羽 真

山口大学 国際総合科学部 講師
博士（文学）

謝辞

本研究は、日本学術振興会科学研究費補助金・
新学術領域研究（研究領域提案型）の課題
「対話知能システムの研究開発及び社会実装の
ための法社会規範の研究」
（研究代表者：新保史生，領域代表者：石黒浩，
課題番号：19H05694）
に基づくものです。

発表目的

発表目的: 哲学・倫理学とロボティクスの新しい協働の形として、「クリティカル・ロボティクス」を提案

- 「クリティカル・ロボティクス」…社会に受け入れられた考え方（価値観）に対する問題提起を試みるロボティクスのアプローチ (Kureha 2023)



主張: クリティカル・ロボティクスは、社会に有益な技術を生み出すことを目的とする「構築的技術哲学」の有望なアプローチたりうる

構築的技術哲学

Brey (2016): 従来の「反省的技術哲学」に対し、「構築的技術哲学」が必要

- 「反省的技術哲学」…技術およびその社会との関係の分析・理解を目指す
- 「構築的技術哲学」…技術の開発・規制・利用に携わるアクターと協働しつつ、社会に有益な技術を生み出すことを目指す
 - 「工学指向型」、「政策指向型」、「利用指向型」の3タイプ

von Schomberg (2019): 技術の設計において「制約に関する倫理」から「構築に関する倫理」への移行が必要

- 「責任ある研究・イノベーション (RRI)」の枠組み内で、倫理(学)の役割を、より社会的に望ましい価値を埋め込んだ技術の構築に見出す

クリティカル・ロボティクスは、従来のELSI・RRIと異なる仕方での、(工学指向型)構築的技術哲学の実践

1. 「クリティカル・ロボティクス」 とは何か?

クリティカル・ロボティクス

従来のロボティクスの目的 (e.g. 石黒 2021):

ロボットの開発を通して…

① 技術的側面: **社会の役に立つものを作る**

➤ 社会に受け入れられた価値観を前提

② 科学的側面: (構成論的アプローチを通して) **人間を理解する**

+

③ **社会に受け入れられた価値観に対して問題を提起する**

➡ 「クリティカル・ロボティクス」

➤ 例)

「弱いロボット」(岡田 2012)、

ロボットのジェンダー表象の研究 (Song-Nichols 2020)、

ロボット謝罪の研究 (Kureha 2023)

クリティカル・ロボティクスの例① —弱いロボット—

「弱いロボット」(岡田 2012)

- 例) 自分でゴミを拾えないゴミ箱ロボット
 - 落ちているものを見つけるとそれを周囲の人に教える仕草をする。それを見た人はゴミを拾ってあげたくなる



- 「1人でできる強さ」を重視する社会への問題提起



クリティカル・ロボティクスの例② —ロボットのジェンダー表象に関する研究—

背景: 女性型のAI・ロボットを作ったり描いたりする際に、ジェンダー・ステレオタイプを強化しかねない仕方でAI・ロボットが描かれてしまう (e.g. Robertson 2017; UNESCO 2019)

- 例) 女性にケア労働を押し付けるような性別役割分業意識

対応 (e.g. WeBel, et al. 2021):

- AI・ロボットへのジェンダー特徴の付与を避ける (「中性化」)
- AI・ロボットにステレオタイプの発言に対して抗議させる
- AI・ロボットにステレオタイプに反する言動を行わせる
- 等

クリティカル・ロボティクスの例② —ロボットのジェンダー表象に関する研究—

例) Song-Nichols (2020):

- 6~8歳の子どもに、反ステレオタイプ的なロボット (例: 女性型の建設労働者ロボット) の動画、または、ステレオタイプ的なロボット (例: 女性型の秘書ロボット) の動画を見せる
- 子どものジェンダーステレオタイプینگが、反ステレオタイプ条件では減少し、ステレオタイプ条件では増加した

- AI・ロボットにステレオタイプに反する言動を行わせる
- 等

ロボットを用いて、社会に蔓延した
ジェンダー・ステレオタイプを是正可能?

クリティカル・ロボティクスの例③ —ロボット謝罪に関する研究 (Kureha 2023)—

背景: 近年、謝罪機能を実装したロボットが登場
(e.g., Uchida, et al. 2019; Okada, et al. 2023)

- 人間との円滑な相互作用を可能にする、と言われる
- しかし、自らの振舞いに責任を負えないロボットに謝罪させることは、人間の責任を曖昧にしかねない
 - ➡ 和解/赦しを妨げる、人間による謝罪の重みを失わせる、等の懸念

問題: ロボットに謝罪させることは許されるか?

答え: ケースバイケース

- 人間だって、謝らなくていい (自分に責任がない) 場面で謝りまくっている
 - ➡ 深刻な害のないケースでは、謝罪を巡る人間社会の慣行に照らせば、ロボットに謝罪させてはならないとは言えない

クリティカル・ロボティクスの例③ —ロボット謝罪に関する研究 (Kureha 2023)—

発表者のアイデア: **ロボットを用いてこの慣行自体を問い直せないか?**

- 例) 「謝らないロボット」で、自身に責任がない場面での謝罪がまかり通る文化を批判

さらなる疑問: **この慣行自体に問題があるのに、それを踏襲していいのか?**

答え: ケースバイケース

- 人間だって、謝らなくていい (自分に責任がない) 場面で謝りまくっている
 - ➔ 深刻な害のないケースでは、**謝罪を巡る人間社会の慣行に照らせば**、ロボットに謝罪させてはならないとは言えない

クリティカル・ロボティクス

発表者の狙い:

- ① これまで散発的になされてきた試みを、ロボティクスのアプローチとして、自覚的・体系的に展開
- ② ロボットを用いた実験で、その効果を検証

③ **社会に受け入れられた価値観に対して問題を提起する**
➔ 「クリティカル・ロボティクス」

➤ 例)

- 「弱いロボット」(岡田 2012)、
- ロボットのジェンダー表象の研究 (Song-Nichols 2020)、
- ロボット謝罪の研究 (Kureha 2023)

2. 構築的技術哲学としての クリティカル・ロボティクス

ELSI・RRIを超えて

主張: クリティカル・ロボティクスは、**哲学・倫理学とロボティクスの新しい協働の形**を示す!

• **従来の研究開発への関与形態:**

➤ **「倫理的・法的・社会的含意/課題 (ELSI)」**…

「科学技術の発展に伴って生じうる影響や課題をその社会実装に先立って特定し、それに対する対応を検討する (…)
取り組み」(呉羽他 2018, p. 8)

➤ **「責任ある研究・イノベーション (RRI)」**…

「(科学技術の進歩が社会に適切に埋め込まれるようにするために)
社会の諸アクターとイノベーターが、イノベーションのプロセスと市場化可能なその成果の(倫理的な)受容可能性、持続可能性、社会的な望ましさに関する見解に相互に応答し合う透明で相互作用的なプロセス」(von Schomberg 2014, p. 39)

ELSI・RRIへの懸念

指摘される懸念 (JST/CRDS 2021):

- 「人文・社会科学の研究者が (...) 単に「異議申し立てをする者」としての役割しか果たせていない」(ibid., p. 43)
- 「人文・社会科学が本来持っていると言われる批判的な精神を失って (...) 理工系の研究者に忖度したり、お墨付きを与えたりするだけの連携に留まる」(「ELSIfication」の問題) (ibid.)

発表者の懸念:

- 技術が生み出す問題とされるものは、実際には元々社会に潜んでいた問題が顕在化 or 深刻化する形で生じることが多い
▶ 例) AIのアルゴリズム差別、ロボットのジェンダー表象
- だが、ELSI・RRIが、こうした社会環境そのものがもつ問題に有効に対処できてきたかは疑問

ELSI・RRIの価値への向き合い方の例①

—ムーンショット目標—

例) ムーンショット型研究開発事業・目標I

「2050年までに、人が身体、脳、空間、時間の制約から解放された社会を実現」

- 「倫理的・法的・社会的・経済的 (ELSE) 課題」にも対応 (萩田 2022)
- 「本研究開発プログラムでは、(...) サイバネティック・アバター技術を、社会通念を踏まえながら研究開発を推進していきます」
(URL=<<https://www.jst.go.jp/moonshot/program/goal1/index.html>>, 2023/12/7閲覧, 強調引用者)

不満: 哲学・倫理学は「社会通念」を教える学問ではない!

- おもしろ、「当たり前を疑う」ことが哲学の仕事とされてきた

ELSI・RRIの価値への向き合い方の例② —価値に敏感なデザイン (VSD)—

RRIでは「価値に敏感なデザイン (VSD)」が主要テーマとされる (von Schomberg 2014)

- 「価値に敏感なデザイン (VSD: value-sensitive design)」…
技術の設計に価値 (特に道徳的価値) への配慮を埋め込むアプローチ
 - 「価値」の例) 健康、所有権、プライバシー、バイアス・フリー、
普遍的利用可能性、信頼、自律性、インフォームドコンセント、
説明可能性、アイデンティティ、持続可能性、等 (Friedman &
Kahn 2003)
- 特徴: ステークホルダーが望む価値



ELSI・RRIの価値への向き合い方の例③ —ジェンダー—

ELSI・RRIにおける、ジェンダー関連の問題への対応:

- 例1) 人工知能学会の「表紙問題」(2014) への対応
 - ➔ 倫理委員会設立にも影響 (松尾他 2015)
 - AI・ロボットのジェンダー表象が社会に対してもつインパクトを意識
 - 例2) Horizon 2020における「研究・イノベーションにおけるジェンダー平等の促進」のための取り組みの公募
 - 研究組織の問題 (研究チームのジェンダー・バランスへの配慮、等) が中心
- ↓
- 研究開発のプロダクトやプロセスにおける問題に対処
 - 逆に技術を用いて社会の中の問題に対処しよう、という発想は、明確に見られない

ELSI・RRIとクリティカル・ロボティクス —価値への向き合い方の相違—

従来のELSI・RRIの「価値」への向き合い方:

- 社会に受け入れられた価値観に依拠する
- アクター・ステークホルダー間の協議を通して、様々な人々に受け入れられうる価値観の創出を目指す
- 社会に受け入れられた考え方 (価値観) 自体を問い直す、
という問題意識は明確に打ち出されていない

それをやるのが、クリティカル・ロボティクス

- 「当たり前を疑う」哲学の批判的機能を活用
➡ 人文社会科学 (哲学) の研究者が科学技術 (ロボット) の研究開発において実質的な貢献を果たすことが可能

3. クリティカル・ロボティクスへの 想定される懸念

想定される懸念① —モラリズム—

懸念: クリティカル・ロボティクスは、**技術の設計によって一定の価値観を強制する「モラリズム」**に与するものではないか？

応答:

- クリティカル・ロボティクスの目的は、特定の価値観の強制ではなく、社会に受け入れられた価値観への問題提起
- 技術はしばしば、意図するかしないかに関わらず、価値を埋め込まれている (van der Hoven 2013)
➡ 埋め込まれるべき価値観についての明示的な考察を促す上で、クリティカル・ロボティクスは有効

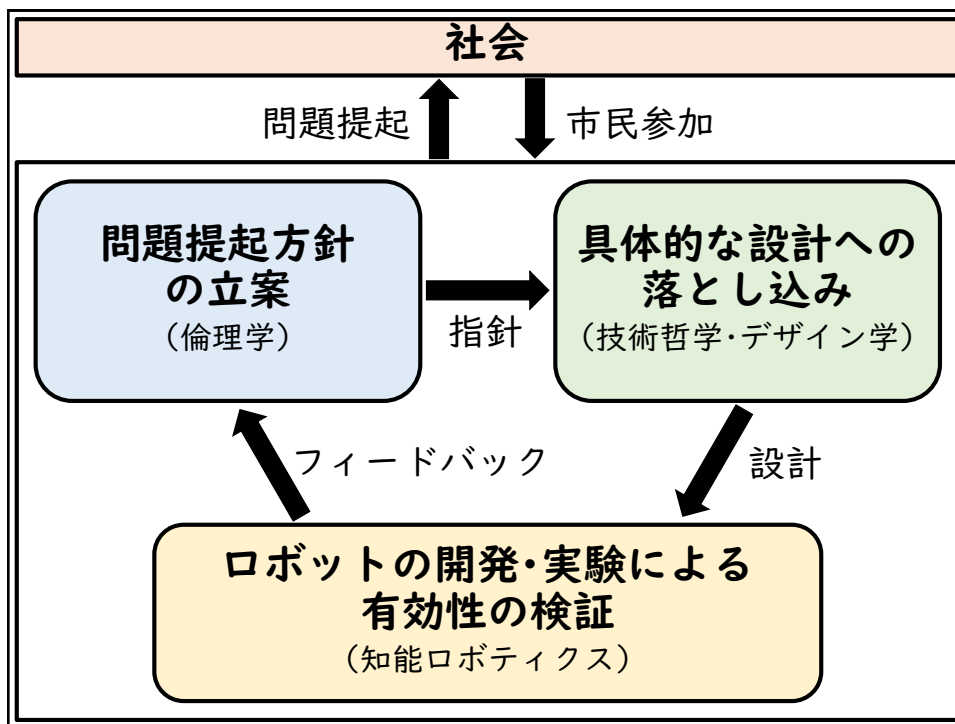
想定される懸念② —悪用—

懸念: クリティカル・ロボティクスのアプローチは、**望ましくない価値観への誘導**に用いられてしまう危険があるのではないか？

- 例) ジェンダー・ステレオタイプ是正の取り組みに対する反動 (バックラッシュ)

応答: クリティカル・ロボティクスのアプローチに、あらかじめ**悪用に対する予防措置**を組み込んでおくことで、一定の対処は可能

- 例) 市民参加型企画 (ワークショップ等) の開催を通して、問題提起の方針に関する意見を募り、設計に反映
- その過程で、誘導に留まらない、問題意識の理解を求める



結論

要約:

- 哲学・倫理学とロボティクスの新しい協働の形として、社会に受け入れられた考え方（価値観）に対する問題提起を行う「クリティカル・ロボティクス」を提案した
- ロボット技術に倫理的かつ批判的な視点を埋め込むクリティカル・ロボティクスは、社会に有益な技術を生み出すことを目指す「構築的技術哲学」の有望なアプローチ
- またこの枠組みでは、人文社会科学の研究者が科学技術の研究開発において実質的な貢献を果たすことが可能

Appendix

クリティカル・デザインの応用としての クリティカル・ロボティクス

「クリティカル・デザイン」(Dunne & Raby 2013)
…技術の設計を通して、社会に受け入れられた先入観や固定観念に揺さぶりをかけることを目指すデザインのアプローチ



ロボットの設計に適用

- ロボットは人間のあり方を映し出す「鏡」(石黒 2009) としての象徴的意味をもつ
 - ➔ 批判的機能を豊かにもつ
 - 「ロボットのふり見てわがふり直せ」



モラル・コンピューティングとの相違

「モラル・コンピューティング」(神田 2021)・・・

人々の行動を道徳的に保ち、安心感をもたらすような
振る舞いをロボットに行わせる計算技術

- 例) スーパーマーケットの挨拶係 ➡ 万引き防止
- 特徴:
 - 道徳性や安心感といった、社会で自明によいとされる価値の実現を目指す
 - もっぱら技術を用いて社会問題を解決しようとする「**技術的問題解決 (technological fix)**」(Weinberg 1966) の一種?

クリティカル・ロボティクスは、技術を用いて慣習的価値観を問い直すことを目指す

文献①

- Brey, P. 2016. 'Constructive philosophy of technology and responsible innovation,' in M. Franssen, P.E. Vermaas, P. Kroes & A.W.M. Meijers, (eds.), *Philosophy of Technology After the Empirical Turn*, pp. 127-143, Springer.
- Dunne, A. & Raby, F. 2013. *Speculative Everything*, MIT Press.
[ダン, A & レイビー, F 2015. 『スペキュラティヴ・デザイン』, 千葉敏生訳, ビー・エヌ・エヌ新社.]
- Friedman, B. & Kahn, Jr., P.H. 2003. 'Human values, ethics and design,' in J.A. Jacko & A. Sears (eds.), *Handbook of Human-Computer Interaction*, pp. 1177-1201, Lawrence Erlbaum.
- Kureha, M. 2023. 'On the moral permissibility of robot apologies,' *AI & Society*, <https://doi.org/10.1007/s00146-023-01782-2>
- Okada, Y., Kimoto, M., Iio, T., Shimohara, K. & Shiomi, M. 2023. 'Two is better than one,' *PLoS ONE* 18: e0281604.
- Robertson, J. 2017. *Robo Sapiens Japonicus*, University of California Press.
- Song-Nichols, K. & Young, A. G. 2020. 'Gendered robots can change children's gender stereotyping,' in *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society (CogSci 2020)*, pp. 2480-2485.

文献②

- Uchida, T., Minato, T., Koyama, T. & Ishiguro, H. 2019. 'Who is responsible for a dialogue breakdown?' *Frontiers in Robotics & AI* 6: 29.
- UNESCO 2019. *I'd Blush If I Could*. URL=<<https://en.unesco.org/Id-blush-if-i-could>>, 2023/12/7閲覧.
- van den Hoven, J. 2013. 'Value-sensitive design and responsible innovation,' in R. Owen, J. Bessant, & M. Heintz (eds.), *Responsible Innovation*, pp. 75-83, Wiley.
- von Schomberg, R. 2014. 'The quest for the 'right' impacts of science and technology,' in J. van den Hoven, et al., (eds.), *Responsible Innovation I*, pp. 33-50, Springer.
- ——— 2019. 'Why responsible innovation,' in R. von Schomberg & J. Hankins (eds.), *International Handbook on Responsible Innovation*, pp. 12-32, Edward Elgar Publishing.
- Weinberg, A.M. 1966. 'Can technology replace social engineering?' *Bulletin of the Atomic Scientists* 12: 4-8.
- WeBel, M., Ellerich-Groppe, N. & Schweda, M. 2021. 'Gender stereotyping of robotic systems in eldercare,' *International Journal of Social Robotics*, <https://doi.org/10.1007/s12369-021-00854-x>

文献③

- 石黒浩 2021. 『ロボットと人間』, 岩波書店.
- 岡田美智男 2012. 『弱いロボット』, 医学書院.
- 科学技術振興機構研究開発戦略センター (JST/CRDS) 2021. 『ELSIからRRIへの展開から考える科学技術・イノベーションの変革』 (CRDS-FY2021-RR-07). URL=<<https://www.jst.go.jp/crds/pdf/2021/RR/CRDS-FY2021-RR-07.pdf>>, 2023/12/7閲覧.
- 神田崇行 2021. 「モラルコンピューティングの研究開発」, 『人工知能』 36: 564-569.
- 呉羽真, 伊勢田哲治, 磯部洋明, 大庭弘継, 近藤圭介, 杉本俊介, 玉澤春史 2018. 『将来の宇宙探査・開発・利用がもつ倫理的・法的・社会的含意に関する研究調査報告書』 . URL=<<https://www.usss.kyoto-u.ac.jp/wp-content/uploads/2021/02/booklet.pdf>>, 2023/12/7閲覧.
- 萩田紀博 2022. 「メタバース動向から見たムーンショット目標Iにおける社会・経済・法・倫理の課題」, 『情報通信政策研究』 6: IA-61-IA-74.
- 松尾豊, 西田豊明, 堀浩一, 武田英明, 長谷敏司, 塩野誠, 服部宏充 2015. 「人工知能学会倫理委員会の取組み」, 『人工知能』 30: 358-364.