

AI 原則は機能するか？

AI・ロボットを用いることに伴う原則策定の取り組みから
法定事項としての位置づけへ

慶應義塾大学 新保 史生

情報通信法学研究会 AI分科会

(令和元年度第1回)

令和元年12月19日(木)13:00-16:00

2020年はAI原則公表ブームが到来する(予言)

- 総務省「AI開発原則(主にAIネットワーク化の健全な進展及びAIシステムの便益の増進に関する原則)」及び「AI利活用原則」を踏襲した原則の策定が行われるようになる
- 2019年末の時点で国内では10社程度がAI原則を公表している状況
- 2020年は、AI原則ブームにより数多くの事業者や組織等が原則を公開することが予想される
- 研究においても、どのようなAI原則が公表され、それらを構成する原則の内容について調査や確認がなされる機会が増えると予想される

本来のAI原則の意図や意義とは異なる方向に進む懸念

- 形式的な原則公開
- コピー原則の公開による原則の意義の希釈化
- 個人情報保護方針の公表と同様の問題が生ずるおそれ

原則を公表する意義

- 原則は、その原則が公表されていることに意味があるのではなく、その原則が必要であるから原則となっているのであって、その原則を踏まえた取り組みをすることが目的のはず
- 原則の原則による原則のための原則が策定され公表されることが目的ではなく、なぜその原則が必要でその原則に基づいて何をすべきなのかを考える段階に来ている

本報告の目的（以下のようなブレインストーミングを行うこと）

AI・ロボットを用いることに伴う原則策定の取り組み

試行錯誤

検討・研究・開発段階

法定事項としての位置づけへ（**法定の原則**）

法規範

実用・実装段階

2015年の「ロボット法 新8原則（新保試案）」の着想から

2020年の「新興技術活用推進基本法（通称「エマテク新法」（新保試案））」の提案へ

本法律はイメージです

高度情報通信ネットワーク社会形成基本法を
（コピー）参考にした立法イメージ

基本理念

すべての国民が新興技術の恵沢を享受できる社会の実現

経済構造改革の推進及び産業国際競争力の強化

ゆとりと豊かさを実感できる国民生活の実現

活力ある地域社会の実現及び住民福祉の向上

国及び地方公共団体と民間との役割分担

利用の機会等の格差の是正

社会経済構造の変化に伴う新たな課題への対応

法制上の措置等

統計等の作成及び公表

国民の理解を深めるための措置

施策の策定に係る基本方針

新興技術の一層の拡充等の一体的な推進

世界最高水準の新興技術の研究開発

教育及び学習の振興並びに人材の育成

新興技術活用取引等の促進

行政における新興技術活用の推進

公共分野における新興技術の活用

新興技術の安全性の確保等

研究開発の推進

国際的な協調及び貢献

AI・ロボット関係 原則・ガイドライン策定の現況

参考文献

- Anna Jobin, Marcello Lenca, Effy Vayena, Artificial Intelligence: the global landscape of ethics guidelines, Nat. Mach. Intell. (2019)
- Marc Rotenberg, The EPIC AI Policy Sourcebook, Electronic Privacy Information Center (EPIC) (September 7, 2019)
- 栗原 佑介「Society5.0時代における企業のAI倫理ガイドライン ～NTT データへのインタビュー含む」InfoCom T&S World Trend Report、2019年11月号(通巻367号) PP.26-33。
- Roger Clarke, Principles for AI: A SourceBook(Revision of 15 April 2019)<<http://www.rogerclarke.com/EC/GAIP.html>>

公的部門

| | 名称 | 原題 | 作成主体 | 分野 | 作成国・地域 | 発行年 | URL |
|----|--|---|---|--------|--------|------|--|
| 1 | 欧州における新興ロボット技術の規制: ロボティクスにおける法と倫理 | Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics, FP7-SCIENCE-IN-SOCIETY-2011-1, Project No.: 289092 | FP7の社会における科学プロジェクト | 政府機関 | EU | 2014 | https://cordis.europa.eu/project/rcn/102044/en |
| 2 | 人工知能の未来への準備 欧州における新興ロボット技術の規制: ロボティクスにおける法と倫理 | Preparing for the future of Artificial Intelligence Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics, FP7-SCIENCE-IN-SOCIETY-2011-1, Project No.: 289092 | Executive Office of the President; National Science and Technology Council, Committee on Technology FP7の社会における科学プロジェクト | 政府機関 | 米国 | 2016 | https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf https://cordis.europa.eu/project/rcn/102044/en |
| 3 | 国立人工知能研究開発戦略計画 | The National Artificial Intelligence Research and Development Strategic Plan | National Science and Technology Council; Networking and Information Technology Research and Development Subcommittee | 政府機関 | 米国 | 2016 | https://www.nitrd.gov/PUBS/national_ai_rd_strategic_plan.pdf |
| 4 | 「人工知能と人間社会に関する懇談会」報告書 | | 内閣府人工知能と人間社会に関する懇談会 | 政府機関 | 日本 | 2017 | https://www8.cao.go.jp/cstp/tyousakai/ai/summary/aisociety_jp.pdf |
| 5 | 国際的な議論のためのAI 開発ガイドライン「AI開発原則」 | | 総務省AI ネットワーク社会推進会議 | 政府機関 | 日本 | 2017 | http://www.soumu.go.jp/menu_news/s-news/01icp01_02000067.html |
| 6 | 欧州評議会、欧州議会勧告2102 (2017)「技術的収束、人工知能及び人権」 | Recommendation 2102 (2017) 1, Technological convergence, artificial intelligence and human rights | 欧州評議会議員会議 | 政府機関 | EU | 2017 | http://www.assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=24236&lang=EN |
| 7 | ロボティクスに係る民事法的規則に関する欧州委員会への提言 | Report with recommendations to the Commission on Civil Law Rules on Robotics | 欧州議会 | 立法機関 | EU | 2017 | http://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html |
| 8 | ビッグデータ、人工知能、機械学習、データ保護 | Big data, artificial intelligence, machine learning and data protection | ICO | 独立政府機関 | イギリス | 2017 | https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf |
| 9 | 知的情報社会に向けた中長期マスタープラン | Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society | 大韓民国政府 | 政府機関 | 韓国 | 2017 | http://english.msip.go.kr/english/msip/Contents/contentsView.do?catelId=msse56&artId=1352869 |
| 10 | 人間はいかに優位性を保つことができるか? AIアルゴリズムにより提起された倫理的問題に関するレポート | How can humans keep the upper hand? Report on the ethical matters raised by AI algorithms | CNIL | 独立政府機関 | フランス | 2017 | https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf |

公的部門

| | 名称 | 原題 | 作成主体 | 分野 | 作成国・地域 | 発行年 | URL |
|----|--|--|--|---------|----------|------|---|
| 11 | AI、ロボット、自律型システムに関するステートメント | Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems | 欧州委員会、科学および新技術における倫理に関する欧州グループ European Commission, European Group on Ethics in Science and New Technologies | 政府機関 | EU | 2018 | http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf |
| 12 | 人工知能の未来のためのシャルルボワの共通のビジョン | Charlevoix Common Vision for the Future of Artificial Intelligence | G7 Leaders of the G7 | 政府間国際組織 | 国際機関・組織等 | 2018 | https://www.mofa.go.jp/files/000373837.pdf |
| 13 | 人工知能における倫理とデータ保護に関する宣言 | Declaration on ethics and data protection in Artificial Intelligence | ICDPPC ICDPPC | 政府間国際組織 | 国際機関・組織等 | 2018 | http://www.amchameu.eu/system/files/position_papers/amchameu_icdppc_declaration_on_ai_final.pdf |
| 14 | 人工知能(AI)コミュニケーションに関する調整計画 2018 | Coordinated plan on Artificial Intelligence (AI) Communication 2018 | 欧州連合、欧州委員会 European Union, European Commission | 政府機関 | EU | 2018 | https://ec.europa.eu/knowledge4policy/publication/coordinated-plan-artificial-intelligence-com2018-795-final_en |
| 15 | AI原則と倫理 | AI Principles & Ethics | スマート・ドバイ Smart Dubai | 政府機関 | アラブ首長国連邦 | 2018 | https://www.smartdubai.ae/pdfviewer/web/viewer.html?file=https://www.smartdubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf?sfvrsn=d4184f8d_6 |
| 16 | 英国におけるAI: 準備万端で活用可能か? | AI in the UK: ready, willing and able? | 英国下院、人工知能に関する特別委員会 UK House of Lords, Select Committee on Artificial Intelligence | 立法機関 | イギリス | 2018 | https://publications.parliament.uk/pa/ld201719/ldselect/lda/100/100.pdf |
| 17 | ディスカッションペーパー: 人工知能の国家戦略 | Discussion Paper: National Strategy for Artificial Intelligence | National Institution for Transforming India (Niti Aayog) National Institution for Transforming India (Niti Aayog) | 政府機関 | インド | 2018 | https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf |
| 18 | 人工知能(AI)と個人データに関するディスカッションペーパー-責任ある開発とAIの採用の促進 | Discussion Paper on Artificial Intelligence (AI) and Personal Data - Fostering Responsible Development and Adoption of AI | 個人情報保護委員会 Personal Data Protection Commission Singapore | 独立政府機関 | シンガポール | 2018 | https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD--050618.pdf |
| 19 | 人工知能とプライバシー | Artificial intelligence and privacy | ノルウェー・データ保護局 The Norwegian Data Protection Authority | 独立政府機関 | ノルウェー | 2018 | https://www.datatilsynet.no/globalassets/global/english/ai-and-privacy.pdf |
| 20 | 欧州委員会から欧州議会、欧州理事会、欧州理事会、欧州経済社会委員会及び欧州人工知能地域委員会への連絡事項 | Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe | 欧州連合、欧州委員会 European Union, European Commission | 政府機関 | EU | 2019 | https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018AE5386 |
| 21 | ロボット倫理に関するCOMESTの報告 | Report of COMEST on Robotics Ethics | COMEST/UNESCO COMEST/UNESCO | 政府間国際組織 | 国際機関・組織等 | 2018 | http://www.academyforlife.va/content/dam/pav/documenti%20pdf/2019/Assamblea2019/Documentazione/01_COMEST%20Report.pdf |
| 22 | AI活用ガイドライン「AI利活用原則」 | | 総務省AI ネットワーク社会推進会議 | 政府機関 | 日本 | 2019 | http://www.soumu.go.jp/menu_news/s-news/01iicp01_0200081.html |
| 23 | 人間中心のAI 社会原則 | | 人間中心のAI社会原則検討会議 - 総合科学技術・イノベーション会議 | 政府機関 | 日本 | 2019 | https://www8.cao.go.jp/cstp/aigensoku.pdf |
| 24 | 信頼できるAIに関する倫理ガイドライン | Ethics Guidelines for Trustworthy AI | 人工知能に関する高等専門家グループ High-Level Expert Group on Artificial Intelligence | 公的機関 | EU | 2019 | https://ec.europa.eu/futurium/en/ai-alliance-consultation |

公的部門における個別分野

| | 名称 | 原題 | 作成主体 | 分野 | 作成国・地域 | 発行年 | URL |
|----|---|---|--|------|----------|------|---|
| 25 | 自動運転とコネクテッドドライビング: レポート | Automated and Connected Driving: Report | 連邦運輸デジタルインフラ省、倫理委員会 | 政府機関 | ドイツ | 2017 | https://www.bmvi.de/SharedDocs/EN/Documents/G/ethic-commission-report.pdf?__blob=publicationFile |
| 26 | 司法制度及び当該環境における人工知能の利用に関する欧州倫理憲章 | European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment | 欧州評議会: 欧州司法制度効率化委員会 (CEPEJ) | 政府機関 | EU | 2018 | https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c |
| 27 | アルゴリズムによる正義についての勧告提案: 警察及び刑事司法制度における人工知能の役割 | Motion for a recommendation about Justice by algorithm . the role of artificial intelligence in policing and criminal justice systems | 欧州評議会議員会議 | 政府機関 | EU | 2018 | http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-en.asp?fileid=25062&lang=en |
| 28 | ヘルスケアH-480.940の拡張インテリジェンスに関するポリシーの推奨事項 | Policy Recommendations on Augmented Intelligence in Health Care H-480.940 | American Medical Association (AMA) | 政府機関 | 米国 | 2018 | https://www.nitrd.gov/rfi/ai/2018/AI-RFI-Response-2018-James-Madara-AMA.pdf |
| 29 | 人工知能時代における労働: 経済、雇用、スキル、倫理に関する4つの視点 | Work in the age of artificial intelligence. Four perspectives on the economy, employment, skills and ethics | 経済雇用省 | 政府機関 | フィンランド | 2018 | http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160980/T/EMjul_21_2018_Work_in_the_age.pdf |
| 30 | 人工知能: ジェンダーインクルージョンに関する未解決の問題 | Artificial Intelligence: open questions about gender inclusion | W20 | 公的機関 | 国際機関・組織等 | 2018 | http://webfoundation.org/docs/2018/06/AI-Gender.pdf |
| 31 | データ駆動型のヘルスケア技術において最初に取り組むべき行動規範 | Initial code of conduct for data-driven health and care technology | 英国保健社会福祉省 | 政府機関 | イギリス | 2019 | https://www.gov.uk/government/publications/code-of-conduct-for-data-driven-health-and-care-technology/initial-code-of-conduct-for-data-driven-health-and-care-technology |
| 32 | 放射線学におけるAIの倫理: 欧州および北米の多社会声明 | Ethics of AI in Radiology: European and North American Multisociety Statement | American College of Radiology; European Society of Radiology; Radiology Society of North America; Society for Imaging Informatics in Medicine; European Society of Medical Imaging Informatics; Canadian Association of Radiologists; American Association of Physicists in Medicine | 公的機関 | 国際機関・組織等 | 2019 | https://link.springer.com/content/pdf/10.1186%2Fs13244-019-0785-8.pdf |

民間部門

| | 名称 | 原題 | 作成主体 | | 分野 | 作成国・地域 | 発行年 | URL |
|----|---|--|--|--|--------|----------|------|---|
| 33 | ロボティクス原則 | Principles of robotics | 英国工学及び物理科学研究評議会 (EPSRC) | Engineering and Physical Sciences Research Council UK (EPSRC) | 学術団体 | イギリス | 2011 | https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/ |
| 34 | ロボットの倫理 | Ethique de la recherche en robotique | CERNA (アリストエン) | CERNA (Allistene) | その他 | フランス | 2014 | http://allistene.org/digitalAssets/38/38704_Avis_robotique_livret.pdf |
| 35 | ビッグデータ分析のための統合倫理フレーム。IAFビッグデータ倫理イニシアチブ、パートA | Unified Ethical Frame for Big Data Analysis. IAF Big Data Ethics Initiative, Part A | The Information Accountability Foundation | The Information Accountability Foundation | その他 | イギリス | 2015 | http://informationaccountability.org/wp-content/uploads/IAF-Unified-Ethical-Frame.pdf |
| 36 | AI Nowレポート: 短期的な人工知能技術の社会的および経済的洞察 | The AI Now Report. The Social and Economic Implications of Artificial Intelligence Technologies in the Near-Term | AI Now Institute | AI Now Institute | 教育研究機関 | 米国 | 2016 | https://ainowinstitute.org/AI_Now_2016_Report.pdf |
| 37 | 信条 | Tenets | Partnership on AI | Partnership on AI | 民間団体 | 国際機関・組織等 | 2016 | https://www.partnershiponai.org/tenets/ |
| 38 | 責任あるアルゴリズムの原則とアルゴリズムの社会的影響に関する声明 | Principles for Accountable Algorithms and a Social Impact Statement for Algorithms | Fairness, Accountability, and Transparency in Machine Learning (FATML) | Fairness, Accountability, and Transparency in Machine Learning (FATML) | 民間団体 | 国際機関・組織等 | 2016 | http://www.friedler.net/papers/principles.pdf |
| 39 | ロボティクスと人工知能の位置づけ | Position on Robotics and Artificial Intelligence | The Greens (グリーンワーキンググループロボット) | The Greens (Green Working Group Robots) | その他 | EU | 2016 | https://www.academia.edu/38369592/Robotics_artificial_intelligence_and_the_evolving_nature_of_work |
| 40 | 人工知能学会倫理指針 | | 人工知能学会 | | 学術団体 | 日本 | 2017 | http://ai-elsi.org/wp-content/uploads/2017/02/人工知能学会倫理指針.pdf |
| 41 | デジタル決定 | Digital Decisions | Center for Democracy & Technology | Center for Democracy & Technology | NPO | 米国 | 2017 | |
| 42 | AI Now 2017レポート | AI Now 2017 Report | AI Now Institute | AI Now Institute | 教育研究機関 | 米国 | 2017 | https://ainowinstitute.org/AI_Now_2017_Report.pdf |
| 43 | 人工知能と機械学習: ポリシーペーパー | Artificial Intelligence and Machine Learning: Policy Paper | Internet Society | Internet Society | NPO | 国際機関・組織等 | 2017 | https://www.internetsociety.org/wp-content/uploads/2017/08/ISOC-AI-Policy-Paper_2017-04-27_0.pdf |
| 44 | 人工知能とデータ分析の倫理原則 | Ethical Principles for Artificial Intelligence and Data Analytics | ソフトウェアおよび情報産業協会 (SIIA) 公共政策部 | Software & Information Industry Association (SIIA), Public Policy Division | 業界団体 | 国際機関・組織等 | 2017 | http://www.sii.net/Portals/0/ppt/Policy/Ethical%20Principles%20for%20Artificial%20Intelligence%20and%20Data%20Analytics%20SIIA%20Issue%20Brief.pdf?ver=2017-11-06-160346-990 |
| 45 | ITI AIポリシーの原則 | ITI AI Policy Principles | 情報技術産業評議会 (ITI) | Information Technology Industry Council (ITI) | 業界団体 | 国際機関・組織等 | 2017 | https://www.iti.org/public-policy/ITIAIPolicyPrinciplesFINAL.pdf |

民間部門

| | 名称 | 原題 | 作成主体 | | 分野 | 作成国・地域 | 発行年 | URL |
|----|---|---|---|---|-----------|----------|------|---|
| 46 | 倫理的に整合した設計：自律及びインテリジェントシステムが人間の幸福を優先するためのビジョン、バージョン2 | Ethically Aligned Design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, version 2 | Institute of Electrical and Electronics Engineers (IEEE)、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems | Institute of Electrical and Electronics Engineers (IEEE)、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems | 国際規格団体 | 国際機関・組織等 | 2017 | https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf |
| 47 | 倫理的な人工知能のトップ10原則 | Top 10 Principles for Ethical Artificial Intelligence | UNIグローバルユニオン | UNI Global Union | 組合 | 国際機関・組織等 | 2017 | http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf |
| 48 | 倫理的に整合した設計：自律的でインテリジェントなシステムによる人間の幸福を優先するためのビジョン、初版 (EAD1e) | Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition (EAD1e) | Institute of Electrical and Electronics Engineers (IEEE)、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems | Institute of Electrical and Electronics Engineers (IEEE)、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems | 国際規格団体 | 国際機関・組織等 | 2017 | https://ethicsinaction.ieee.org/ |
| 49 | 機械学習：事例から学習するコンピュータの能力と約束 | Machine learning: the power and promise of computers that learn by example | 王立協会 | The Royal Society | その他 | イギリス | 2017 | https://royalsociety.org/-/media/policy/projects/machine-learning/publications/machine-learning-report.pdf |
| 50 | 人工知能の悪意のある利用：予測、防止、および軽減 | The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation | Institute of Electrical and Electronics Engineers (IEEE)、The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems | Future of Humanity Institute; University of Oxford; Centre for the Study of Existential Risk; University of Cambridge; Center for a New American Security; Electronic Frontier Foundation; OpenAI | 民間組織／教育機関 | 国際機関・組織等 | 2018 | https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf |
| 51 | ホワイトペーパー：機械学習で差別的な結果を防ぐ方法 | White Paper: How to Prevent Discriminatory Outcomes in Machine Learning | WEF, Global Future Council on Human Rights 2016-2018 | WEF, Global Future Council on Human Rights 2016-2018 | NPO | 国際機関・組織等 | 2018 | http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf |
| 52 | 人工知能時代におけるプライバシーと表現の自由 | Privacy and Freedom of Expression In the Age of Artificial Intelligence | Privacy International | Privacy International & Article 19 | NPO | 国際機関・組織等 | 2018 | https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf |
| 53 | トロント宣言：機械学習システムにおける平等と差別されない権利の保障 | The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems | アクセス・ナウ；アムネスティ・インターナショナル | Access Now ; Amnesty International | NPO/NGO | 国際機関・組織等 | 2018 | https://www.accessnow.org/cms/assets/uploads/2018/08/The-Toronto-Declaration_ENG_08-2018.pdf |
| 54 | AI4People—より良いAI社会のための倫理的枠組み：機会、リスク、原則、および推奨事項 | AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations | AI4People | AI4People | 有志の集まり | EU | 2018 | https://link.springer.com/article/10.1007/s11023-018-9482-5 |
| 55 | 人工知能のためのユニバーサル・ガイドライン | Universal Guidelines for Artificial Intelligence | The Public Voice | The Public Voice | NGO | 国際機関・組織等 | 2018 | https://epic.org/International/AIGuidelinesDRAFT20180910.pdf |
| 56 | モントリオール宣言：責任あるAI | Montréal Declaration: Responsible AI | モントリオール大学 | Université de Montréal | 教育機関 | カナダ | 2018 | https://ai.quebec/wp-content/uploads/sites/2/2018/12/News-release_Launch_Montreal_Declaration_AI-04_12_18.pdf |
| 57 | ビジネス倫理と人工知能 | Business Ethics and Artificial Intelligence | ビジネス倫理研究所 | Institute of Business Ethics | 民間団体 | イギリス | 2018 | https://www.ibe.org.uk/userassets/briefing/s/ibe_briefing_58_business_ethics_and_artificial_intelligence.pdf |
| 58 | AI原則 | AI Principles | Future of Life Institute | Future of Life Institute | その他 | 米国 | 2018 | |
| 59 | OpenAI憲章 | OpenAI Charter | OpenAI | OpenAI | NPO | 米国 | 2018 | |
| 60 | 人工知能の管理、人権と尊厳の維持 | Governing Artificial Intelligence. Upholding Human Rights & Dignity | Data & Society | Data & Society | NPO | 米国 | 2018 | https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf |

民間部門

| | 名称 | 原題 | 作成主体 | | 分野 | 作成国・地域 | 発行年 | URL |
|----|-------------------------------------|--|--|---|--------|----------|-------|---|
| 61 | 科学、法と社会(SLS)イニシアチブ | Science, Law and Society (SLS) Initiative | The Future Society | The Future Society | NPO | 米国 | 2018 | https://www.institut.veolia.org/sites/g/files/dvc2551/files/document/2018/03/Intervention_Nicolas_Miallhe_-_conf._Institut_27_fevrier_2018.pdf |
| 62 | AI Now 2018レポート | AI Now 2018 Report | AI Now Institute | AI Now Institute | 教育研究機関 | 米国 | 2018 | https://ainowinstitute.org/AI_Now_2018_Report.pdf |
| 63 | ロボット時代の人権報告 | Human Rights in the Robot Age Report | ラテナウ研究所 | The Rathenau Institute | 研究機関 | オランダ | 2018 | https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf |
| 64 | オランダの人工知能マニフェスト | Dutch Artificial Intelligence Manifesto | 人工知能に関する特別グループ(SIGAI)、ICT Platform Netherlands (IPN) | Special Interest Group on Artificial Intelligence (SIGAI), ICT Platform Netherlands (IPN) | その他 | オランダ | 2018 | http://ii.tudelft.nl/bnvki/wp-content/uploads/2018/09/Dutch-AI-Manifesto.pdf |
| 65 | 日本医師会学術推進会議 | | 第IX次学術推進会議報告書人工知能(AI)と医療 | | 民間団体 | 日本 | 2018 | https://www.med.or.jp/nichiionline/article/006805.html |
| 66 | 健康分野における人工知能(AI) | Artificial Intelligence (AI) in Health | 王立医科大学 | Royal College of Physicians | 教育機関 | イギリス | 2018 | https://www.rcplondon.ac.uk/projects/outputs/artificial-intelligence-ai-health |
| 67 | 健康に関する人工知能の倫理的、社会的、政治的課題 | Ethical, social, and political challenges of Artificial Intelligence in Health | Future Advocacy | Future Advocacy | 教育機関 | イギリス | 2018 | https://wellcome.ac.uk/sites/default/files/ai-in-health-ethical-social-political-challenges.pdf |
| 68 | 日本経済団体連合会 | | AI活用戦略～AI-Readyな社会の実現に向けて～ | | 業界団体 | 日本 | 2019 | https://www.keidanren.or.jp/policy/2019/013.html |
| 69 | 責任あるAIの10原則 | 10 Principles of responsible AI | Women leading in AI | Women leading in AI | 民間団体 | 国際機関・組織等 | 2019 | https://womenleadinginai.org/report2019 |
| 70 | 人工知能:オーストラリアの倫理フレームワーク:ディスカッションペーパー | Artificial Intelligence. Australia's Ethics Framework. A discussion Paper | Department of Industry Innovation and Science | Department of Industry Innovation and Science | 有志の集まり | オーストラリア | 2019 | http://datagovernanceaus.com.au/wp-content/uploads/2016/07/Future-AI-Forum-AI-Ethics-Framework-Consultation-Submission-FINAL.pdf |
| 71 | 倫理指針 | Ethics Policy | アイスランド・インテリジェント・マシン研究所 (IIIM) | Icelandic Institute for Intelligent Machines (IIIM) | その他 | アイスランド | 公表年不明 | http://www.iiim.is/ethics-policy/ |
| 72 | ディープマインドの倫理と社会的原則 | DeepMind Ethics & Society Principles | ディープマインドの倫理と社会 | DeepMind Ethics & Society | その他 | イギリス | 公表年不明 | https://deepmind.com/about/ethics-and-society |
| 73 | 倫理フレームワーク-責任あるAI | Ethics Framework - Responsible AI | マシン・インテリジェンス・ガレージ倫理委員会 | Machine Intelligence Garage Ethics Committee | その他 | イギリス | 公表年不明 | https://www.migarage.ai/ethics-framework/ |

民間企業

| | 名称 | 原題 | 作成主体 | | 分野 | 作成国・地域 | 発行年 | URL |
|----|---|--|--------------------------------------|--------------------------------------|------|--------|------------|---|
| 74 | ソニーグループAI倫理ガイドライン | Sony x AI - AI倫理ガイドライン - Sony Japan | ソニー | | 民間企業 | 日本 | 発行 2018 | https://www.sony.co.jp/SonyInfo/csr_repo/2018/rt/humanrights/hkrfmg000000610b-改定att/AI_Engagement_within_Sony_Group_2019Ja.pdf |
| 75 | NECグループ AIと人権に関するポリシー～AI・生体情報などデータ利用時においてプライバシーへの配慮・人権尊重を最優先し事業を推進～ | | NEC | | 民間企業 | 日本 | 2019 | https://jpn.nec.com/press/201904/20190402_01.html |
| 76 | AIの安心・安全な利用に向けた「富士通グループAIコミットメント」 | | 富士通株式会社 | | 民間企業 | 日本 | 2019 | https://pr.fujitsu.com/jp/news/2019/03/13-1.html |
| 77 | NTTドコモ パーソナルデータ憲章 | | NTTドコモ | | 民間企業 | 日本 | 2019 | https://www.nttdocomo.co.jp/info/notice/pages/190827_00.html |
| 78 | NTT データグループAI 指針 | | NTT データ | | 民間企業 | 日本 | 2019 | https://www.nttdata.com/jp/ja/news/releases/2019/052900/ |
| 79 | 倫理規範:5つの基本原則から成るビジネス向けAIの開発 | The Ethics of Code: Developing AI for Business with Five Core Principles | Sage Group pl | Sage Group pl | 民間企業 | イギリス | 2017 | https://www.sage.com/au/resources/enterprise-blog/2017/the-ethics-of-code |
| 80 | 責任あるAIとロボティクス・倫理的枠組み | Responsible AI and robotics. An ethical framework. | アクセンチュアUK | Accenture UK | 民間企業 | イギリス | 公表年不明 | https://www.accenture.com/gb-en/company-responsible-ai-robotics |
| 81 | 責任あるAIフレームワーク | The responsible AI framework | PriceWaterhouseCoopers UK | PriceWaterhouseCoopers UK | 民間企業 | イギリス | 公表年不明 | https://www.pwc.co.uk/services/risk-assurance/insights/accelerating-innovation-through-responsible-ai/responsible-ai-framework.html |
| 82 | 市民のための人工知能 | L'intelligenza artificiale al servizio del cittadino | Agenzia per l'Italia Digitale (AGID) | Agenzia per l'Italia Digitale (AGID) | 民間企業 | イタリア | 2018 | https://ia.italia.it/assets/librobianco.pdf |
| 83 | テレフォニカのAI原則 | AI Principles of Telefónica | テレフォニカ | Telefonica | 民間企業 | スペイン | 公表年不明 | https://www.telefonica.com/en/web/responsible-business/our-commitments/ai-principles |

民間企業

| | 名称 | 原題 | 作成主体 | | 分野 | 作成国・地域 | 発行年 | URL |
|----|--|--|---|---|------|--------|-------|---|
| 84 | AIガイドライン | AI Guidelines | ドイツテレコム | Deutsche Telekom | 民間企業 | ドイツ | 公表年不明 | https://www.telekom.com/en/company/digital-responsibility/details/artificial-intelligence-ai-guideline-524366 |
| 85 | 人工知能に関するSAPの指針 | SAP's guiding principles for artificial intelligence | SAP | SAP | 民間企業 | ドイツ | 2018 | https://www.sap.com/documents/2018/09/940c6047-1c7d-0010-87a3-c30de2fd8ff.html |
| 86 | TietoのAI倫理ガイドライン | Tieto's AI ethics guidelines | ティエト | Tieto | 民間企業 | フィンランド | 2018 | https://www.tieto.com/contentassets/964a20887f764aae944e4f029d05ff51/tieto-s-ai-ethics-guidelines.pdf |
| 87 | コミットメントと原則 | Commitments and principles | OPグループ | OP Group | 民間企業 | フィンランド | 公表年不明 | https://www.op.fi/op-financial-group/corporate-social-responsibility/commitments-and-principles |
| 88 | 意味ある人工知能のために: フランスとヨーロッパの戦略に向けて | For a meaningful Artificial Intelligence. Towards a French and European strategy | ミッションビルラニ | Mission Villani | 民間企業 | フランス | 2018 | https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf |
| 89 | アルゴリズムの透明性と説明責任に関する声明 | Statement on Algorithmic Transparency and Accountability | Association for Computing Machinery (ACM) | Association for Computing Machinery (ACM) | 民間企業 | 米国 | 2017 | https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithmics.pdf |
| 90 | AI - 当社のアプローチ | AI - Our approach | Microsoft | Microsoft | 民間企業 | 米国 | 公表年不明 | https://www.microsoft.com/ja-jp/ai/our-approach-to-ai |
| 91 | 人工知能: 公共政策の機会 | Artificial Intelligence. The Public Policy Opportunity | Intel Corporation | Intel Corporation | 民間企業 | 米国 | 2017 | https://www.intel.ai/solutions/artificial-intelligence-the-public-policy-opportunity/#gs.l51uxn |
| 92 | 信頼と透明性に関するIBMの原則 | IBM's Principles for Trust and Transparency | IBM | IBM | 民間企業 | 米国 | 2018 | https://www.ibm.com/blogs/policy/wp-content/uploads/2018/05/IBM_Principles_OnePage.pdf |
| 93 | 当社の原則 | Our principles | Google | Google | 民間企業 | 米国 | 公表年不明 | https://ai.google/principles |
| 94 | 人工知能のための日常的な倫理: 設計者及び開発者向けの実務ガイド | Everyday Ethics for Artificial Intelligence. A practical guide for designers & developers | IBM | IBM | 民間企業 | 米国 | 2019 | https://www.ibm.com/watson/assets/duo/pdf/everdayethics.pdf |
| 95 | インテルのAIプライバシーポリシーホワイトペーパー: 人工知能の世界における個人のプライバシーとデータの保護 | Intel's AI Privacy Policy White Paper. Protecting individuals' privacy and data in the artificial intelligence world | Intel Corporation | Intel Corporation | 民間企業 | 米国 | 2018 | https://blogs.intel.com/policy/files/2018/10/Intels-AI-Privacy-Policy-White-Paper-2018.pdf |
| 96 | 倫理的AIのためのUnityの指導原則の紹介- Unityブログ | Introducing Unity's Guiding Principles for Ethical AI - Unity Blog | Unity Technologies | Unity Technologies | 民間企業 | 米国 | 2018 | https://blogs.unity3d.com/2018/11/28/introducing-unitys-guiding-principles-for-ethical-ai/ |
| 97 | 責任あるボット: 対話型AIの開発者向けの10のガイドライン | Responsible bots: 10 guidelines for developers of conversational AI | Microsoft | Microsoft | 民間企業 | 米国 | 2018 | https://www.microsoft.com/en-us/research/uploads/prod/2018/11/Bot_Guidelines_Nov_2018.pdf |

1. EUのロボロー・ガイドライン(2014年9月)

- Regulating Emerging Robotic Technologies in Europe: Robotics facing Law and Ethics, FP7-SCIENCE-IN-SOCIETY-2011-1, Project No.: 289092(報告書作成費用:149万7966ユーロ)
- ロボットをめぐる法的課題の体系的な検討を試みた最初の取り組み
- 体系的なガイドラインの提示には至らず、具体的な問題を手がかりに検討に着手せざるを得なかったとしている

具体的な検討課題

- ①ロボット(ドローンなど文字通りのロボット)
- ②自動運転(自動操縦)
- ③モビルスーツ・義足等(装着型、ウェアラブル)
- ④手術・遠隔地・宇宙(リモート操作、遠隔地対応)
- ⑤医療、介護、福祉(医療等分野)
- ⑥災害・レジリエンス

法的課題について

- ①健康、安全、環境、利用者保護のための規制(安心・安全な利用環境の保護)
- ②法的責任 製造物責任(物の製造物責任、情報の製造物責任は認められるか)
- ③知的財産(ロボットそのもの、ロボットが創作したもの)
- ④プライバシー
- ⑤権利能力(エージェント)

ロボット法 新8原則（新保試案）2015

ロボット法の理念・概念

ロボット法原則 (The Robot Law Principles)

①人間第一の原則 Humanity First

②命令服従の原則 Obedience to Order

③秘密保持の原則 Secrecy

④利用制限の原則 Use Limitation

⑤安全保護の原則 Security Safeguards

⑥公開・透明性の原則 Openness & Transparency

⑦個人参加の原則 Individual Participation

⑧責任の原則 Accountability

アイザック・アシモフのSF小説(I, Robot, Isaac Asimov)
「ロボット工学の三原則」(Three Laws of Robotics)

第1条

ロボットは人間に危害を加えてはならない。また、その危険を看過することによって、人間に危害を及ぼしてはならない。

第2条

ロボットは人間にあたえられた命令に服従しなければならない。ただし、あたえられた命令が、第1条に反する場合は、この限りでない。

第3条

ロボットは、前掲第1条および第2条に反するおそれのないかぎり、自己をまもらなければならない。

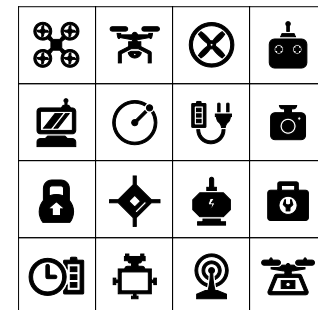
——『ロボット工学ハンドブック』第56版、西暦2058年

半世紀以上前にアシモフが生み出した「思想」であって、「法」や「規範」ではない。

ロボット法 新8原則 (新保試案)

New Eight Principles of Laws of Robotics
(Tentative Proposal by Dr.Shimpo)

OECDプライバシー8原則を参考に
シンギュラリティ(技術的特異点)も見据えて



2015年に「ロボット法 新8原則（新保試案）2015」の着想に至った経緯

- 2015年10月11日に開催した「ロボット法学会設立準備研究会」において、「ロボット法新8原則」を提唱
- 2016年9月10日に開催された「ロボット革命国際シンポジウム（主催：経済産業省、ロボット革命イニシアティブ協議会）」において、「ロボットの社会実装に向けたルール作り(Robot, Society and Rule, Preparation for Robot Installation into Society)」として同8原則を提示

OECDプライバシーガイドラインが1980年に制定された後、各国が個人情報保護制度の整備を行い、現在では世界中の国々が8原則に基づいて法整備を実施していること

- OECDの「プライバシー保護と個人データの流通についてのガイドライン(OECDプライバシーガイドライン)」(Guidelines governing the protection of privacy and transborder flows of personal data)は、1980年9月23日にOECD理事会勧告として採択された。ガイドラインは、OECD加盟国の指針となるものであり、法的拘束力はない。

2013年の当該ガイドライン改正においても、意見対立によって8原則は結局変更されなかったが、8原則を変更せずとも普遍性があることが確認できる結果となったこと

- ガイドラインの改正に向けた議論の過程では、個人の基本的権利保障の枠組みで個人情報・プライバシー保護を実現すべきであるというEUの主張と、クラウドコンピューティングやソーシャルネットワーキングサービス(SNS)に代表されるように、国際的に個人情報が流通する新たなサービスが出現し日々利用されるようになってきている現状を踏まえた検討が必要であるとの意見対立があり、基本的権利としてのプライバシーを根拠に個人情報の取扱いについて絶対的な保障を与えることへの異論も唱えられ、原則の拡充には至らなかった。

原則はあくまで原則でしかないが、普遍性がある原則は法制度において具体的なルールとして組み込むことができること

プライオリティの確保はイニシアティブ獲得につながること

- EUのロボロー・ガイドライン(2014年9月)が公表されたこと

決められたルールを遵守することよりもルールを決めることのほうが難しいこと

Draft SPDE QuestionnaireERRevAug25(非公開資料)

- AI・ロボット原則策定の必要性は、2016年9月9日に実施されたOECDのSPDEビューロ・コールにおいて、エマージング・テクノロジー(IoT、ビッグデータ、AI等)に関する課題において、AIやロボットの利用に伴い基本となる原則の必要性について発言をしたことに端を発する。
- 当該資料の「3.Update on Digital Economy Outlook」の(E2.a. Privacy challenges trends: Currently in your country, what are the 3 most pressing privacy challenges in relation to emerging technologies, (e.g. IoT, big data, artificial intelligence))に係る意見につき、「DEOの質問事項に関して、p1の”E1.a.Digital security challenges trends”の3つの質問の切り口、『経済的活動の課題』、『社会的活動の課題』、『技術上の課題』は漠然としているため、むしろp12の”E2.a.Privacy challenges trends”の3つ「新興技術(IoT、ビッグデータ、AI等)に関する課題』、『政府の取組の課題』、『民間の取組の課題』とすべきであるとの見解を表明、セキュリティも合わせるべきであることと、IoT、ビッグデータ、AI等に係る課題について検討すべきではないかとの見解を提示。
- 当該資料の「4. Discussion of Working Party mandates and working methods」について、「本年4月の日本開催のG7情報通信大臣会合において、IoTやAIなど新たなICTの普及による経済成長の推進について議論され、新たに生じるセキュリティやプライバシーの課題への対応が重要であることが再認識されたことから、SPDEのマネートに関しても、そのような新技術の登場への留意について記載してはどうか。」との意見を表明し、AIやロボットの利用に伴い基本となる原則の必要性についても意見を述べたところ、「p3の3行目のdigital economy の後ろに”,while paying due attention to emergence of new technologies”を追記)」することとなった。
- (その後、2016年11月15日のOECD/SPDE会合において新保はSPDEの副議長を退任)

法的・倫理的側面をめぐる原則等の策定に向けた主な具体的取組みの系譜

総務省「AIネットワーク化検討会議」(旧称:ICTインテリジェント化影響評価検討会議)(2016年1月-6月)、その後、総務省「AIネットワーク社会推進会議」(2016年10月～)

- 「AI開発ガイドライン(2017年7月)」、「AI活用ガイドライン(2018年7月)」

BSI (British Standard Initiative)

- BS 8611:2016, Robots and robotic devices Guide to the ethical design and application of robots and robotic systems(2016年4月)

スタンフォード大学

- AI100「2030年における人工知能と生活」(2016年6月)

Partnership on AI(2016年9月)

- AIの研究・技術について、プライバシーとセキュリティの保護、当事者の利益の理解・尊重、社会的責任、頑健性・堅牢性の確保、人権の尊重、説明可能性などを内容とする「信条」(Tenets)を公表。

英国下院科学技術委員会

- 「ロボティクスと人工知能」(2016年10月12日)

米国ホワイトハウス

- 「人工知能の未来に備えて」(2016年10月)
- AI対応システム(AI-enabled systems)について
 - ①統御可能であること、②オープン、透明性、理解可能であること、③人々と効果的に機能し得ること、④その操作は人間の価値及び願望と一致し続けること

米国電気電子学会(IEEE)

- 「倫理的に調整された設計」(2016年12月)
- 倫理的に調整されたAIの設計に関するIEEEの規格の策定に向けた論点を提示

欧州議会

- 「ロボティクスに係る民事法的規則に関する欧州委員会への提言」(2017年2月)
- ロボット・AIの開発・設計・利用において尊重されるべき基本的な倫理原則を策定するための論点整理
- ロボット・AIを所管するEU機関の設置、スマート・ロボットの登録制、ロボットに関する損害賠償責任、保険、知的財産権、データの流通と保護、標準化等について定めるEU法の策定を提言
 - 夏井高人「ロボット法の制定を求める欧州議会決議 [参考訳]」法と情報雑誌第2巻第5号(通巻第11号・2017年5月号)。
 - その他、EUの関係文書は、法と情報雑誌に夏井高人教授が網羅的に翻訳を掲載しているため、法と情報雑誌を参照のこと。

ITU(国際電気通信連合)

- グローバルサミット「AI for Good」(2017年6月)

AAAI/ACM

- 「AI、倫理、社会に関する会議」(2017年)

中国国務院

- 「次世代人工知能発展計画」(2017年7月)

Future of Life Institute (FLI)

- 「アシロマAI原則」(2017年1月)

内閣府(総合科学技術・イノベーション会議)

- 「人工知能と人間社会に関する懇談会」(2017)

人工知能学会倫理委員会

- 「人工知能学会 倫理指針」(2017年2月28日)

内閣府(総合科学技術・イノベーション会議)人間中心のAI社会原則検討会議

- 「人間中心のAI社会原則」(2018)

英国上院AI特別委員会

- AI報告書(2018年6月)

International Conference of Data Protection and Privacy Commissioners(ICDPPC)

- Declaration on Ethics and Data Protection in Artificial Intelligence(2018年10月)

OECD

- 「AIに関する理事会勧告」(2019年5月22日)

欧州連合「犯罪予防、調査、犯罪捜査目的若しくは犯罪訴追手続又は刑事罰の執行のための主務執行機関による個人データの処理に関する自然人の保護及び当該データの自由な移動並びに欧州議会枠組決定2008/977 / JHAの撤回に関する欧州議会及び欧州理事会2016年4月27日指令2016/680」(2016)

- (European Union (2016) Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA)

欧州評議会議員会議、欧州議会勧告2102(2017)「技術的収束、人工知能及び人権」(2017)

- (Council of Europe, Parliamentary Assembly, Recommendation 2102 (2017) 1, Tech-nological convergence, artificial intelligence and human rights)

欧州連合、欧州委員会「人工知能(AI)コミュニケーションに関する調整計画」

- (European Union, European Commission (2018) Coordinated plan on Artificial Intelligence (AI) Communication)

欧州委員会「欧州委員会から欧州議会、欧州理事会、欧州理事会、欧州経済社会委員会及び欧州人工知能地域委員会への連絡事項」(2018)

- (European Commission (2018) Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe)

欧州評議会「個人データの処理に関する個人保護のための近代化条約」(2018)

- Council of Europe, Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data)

欧州連合「AI、ロボット、自律型システムに関するステートメント」(2018)

- (European Group on Ethics in Science and New Technologies, Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, Brussels, 9 March 2018)

欧州評議会、欧州司法制度効率化委員会「司法制度における人工知能の利用に関する欧州倫理憲章」(2018年)

- (Council of Europe, European Commission for the Efficiency of Justice (2018) European Ethical Charter on the use of artificial intelligence in judicial systems)

欧州評議会議員会議「アルゴリズムによる正義についての勧告提案：警察及び刑事司法制度における人工知能の役割」(2018)

- (Council of Europe, Parliamentary Assembly (2018) Motion for a recommendation about Justice by algorithm . the role of artificial intelligence in policing and criminal justice systems)

欧州連合、欧州委員会の人工知能に関する高等専門家グループ「信頼できるAIに関する倫理ガイドライン」(2019)

- (European Union, European Commission's High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI)

EU「AI、ロボット、自律型システムに関するステートメント」

(European Group on Ethics in Science and New Technologies, Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems, Brussels, 9 March 2018)

概要

- 人工知能、ロボット技術及びいわゆる「自律型」の技術進歩
 - 道徳に係る問題について、喫緊に対応が必要であり、かつ、複雑な問題を提起
- 倫理的、社会的および法的な課題に対する答えを見つけこと
 - 異なるイニシアチブのパッチワークとして共通の利益を求める方向性を示すこと
- 集合的、広範囲、包括的な検討と対話の必要性を強調
 - 社会を組織する上で求められる価値観、社会においてテクノロジーが果たすべき役割に焦点を当てた対話が必要

本文書の公開目的

- 共通の国際的な倫理的及び法的枠組みの構築を目指す取り組みへの着手を示すもの
 - 人工知能、ロボット技術、「自律」システムの設計、製造、使用および管理が対象
- 基本的な倫理原則を提案
 - EUの各条約及びEU基本権憲章に定められた価値観に基づくもの

(1) 安全性、セキュリティ、危害の防止及びリスクの軽減に関する問題

- 相互接続されたAIと「自律型」のデバイスを使用する世界を安全かつ安全にするための方策

(2) 人間の道徳的責任についての問題

- 高度なAIとロボットの構成要素を備え、動的で複雑な社会技術システムについて、道徳的な問題について検討する組織は？
- 道徳的責任の帰属先、不都合な結果についても責任分配
- 「管理の共有(shared control)」や人間とスマートマシンとの関係について検討することは意義があるか
- 道徳的「緩衝領域(crumple zones)」として、人間は「自律的な」装置のエコシステムの一部でしかないのか

(3) ガバナンス、規制、設計、開発、検査、監視、テスト及び認証に関する問題

- 個人や社会の福祉に役立つように、そして社会をこの技術に対して安全にするために、私たちの制度や法律はどのように再設計されるべきか？

(4) 上記の問題の基調となる制度、政策及び価値観に関する意思決定を含む民主的な意思決定の問題

- プロファイリング、マイクロターゲティングを可能にする機械学習、ビッグデータ、行動科学の組み合わせに基づく高度なナッジ技術の使用により市民がどの程度恩恵を受けるのかについて確認するための調査。商業的又は政治的な目的に応じた選択アーキテクチャの調整および操作。

(5) AIと「自律型」システムの説明可能性と透明性についての問題

IEEE's (Institute of Electrical and Electronics Engineers) policy paper on 'Ethically Aligned Design'

• http://standards.ieee.org/news/2016/ethically_aligned_design.html

ITU's (International Telecommunication Union) Global Summit 'AI for Good'

• <https://www.itu.int/en/ITU-T/AI/Pages/201706-default.aspx>

ACM's (Association for Computing Machinery) work on the issue, including a major AAAI/ACM 'Conference on AI, Ethics, and Society' in summer 2017

• <http://www.aies-conference.com/>

IBM, Microsoft and Google's DeepMind have established their own ethic codes on AI and joined forces in creating broad initiatives such as the 'Partnership on AI' in 2018

• <https://www.partnershiponai.org/>

OpenAI

• <https://openai.com/>

Asilomar AI Principles

• <https://futureoflife.org/ai-principles/>

The Forum on the Socially Responsible Development of Artificial Intelligence held by the University of Montreal in November 2017, 'Declaration for a Responsible Development of Artificial Intelligence' has been developed. It is now publicly accessible on an online platform where all sectors of society are invited to comment on the text.

• <http://nouvelles.umontreal.ca/en/article/2017/11/03/montreal-declaration-for-a-responsible-development-of-artificial-intelligence/>

(a) 人間の尊厳 (Human dignity)

(b) 自治 (Autonomy)

(c) 責任 (Responsibility)

(d) 正義、公平及び相互依存 (Justice, equity, and solidarity)


(e) 民主主義 (Democracy)

(f) 法の支配と説明責任 (Rule of law and accountability)

(g) 安全性、安全性、身体的及び精神的完全性 (Security, safety, bodily and mental integrity)

(h) データ保護とプライバシー (Data protection and privacy)

(i) 持続可能性 (Sustainability)



AI と倫理・原則策定の例

- 人工知能学会において、人工知能と社会との関わりについて検討を行い社会に発信することを目的に、2014年12月に倫理委員会が発足

■ 倫理指針を策定

- ① 人類への貢献
- ② 法規制の遵守
- ③ 他者のプライバシーの尊重
- ④ 公正性
- ⑤ 安全性
- ⑥ 誠実な振る舞い
- ⑦ 社会に対する責任
- ⑧ 社会との対話と自己研鑽
- ⑨ 人工知能への倫理遵守の要請

1（人類への貢献）

- 人工知能学会会員は、人類の平和、安全、福祉、公共の利益に貢献し、基本的人権と尊厳を守り、文化の多様性を尊重する。人工知能学会会員は人工知能を設計、開発、運用する際には専門家として人類の安全への脅威を排除するように努める。

2（法規制の遵守）

- 人工知能学会会員は専門家として、研究開発に関わる法規制、知的財産、他者との契約や合意を尊重しなければならない。人工知能学会会員は他者の情報や財産の侵害や損失といった危害を加えてはならず、直接的のみならず間接的にも他者に危害を加えるような意図をもって人工知能を利用しない。

3（他者のプライバシーの尊重）

- 人工知能学会会員は、人工知能の利用および開発において、他者のプライバシーを尊重し、関連する法規に則って個人情報 の適正な取扱いを行う義務を負う。

4（公正性）

- 人工知能学会会員は、人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人工知能学会会員は人類が公平、平等に人工知能を利用できるように努める。

5（安全性）

- 人工知能学会会員は専門家として、人工知能の安全性及びその制御における責任を認識し、人工知能の開発と利用において常に安全性と制御可能性、必要とされる機密性について留意し、同時に人工知能を利用する者に対し適切な情報提供と注意喚起を行うように努める。

6（誠実な振る舞い）

- 人工知能学会会員は、人工知能が社会へ与える影響が大きいことを認識し、社会に対して誠実に信頼されるように振る舞う。人工知能学会会員は専門家として虚偽や不明瞭な主張を行わず、研究開発を行った人工知能の技術的限界や問題点について科学的に真摯に説明を行う。

7（社会に対する責任）

- 人工知能学会会員は、研究開発を行った人工知能がもたらす結果について検証し、潜在的な危険性については社会に対して警鐘を鳴らさなければならない。人工知能学会会員は意図に反して研究開発が他者に危害を加える用途に利用される可能性があることを認識し、悪用されることを防止する措置を講じるように努める。また、同時に人工知能が悪用されることを発見した者や告発した者が不利益を被るようなことがないように努める。

8（社会との対話と自己研鑽）

- 人工知能学会会員は、人工知能に関する社会的な理解が深まるよう努める。人工知能学会会員は、社会には様々な声があることを理解し、社会から真摯に学び、理解を深め、社会との不断の対話を通じて専門家として人間社会の平和と幸福に貢献することとする。人工知能学会会員は高度な専門家として絶え間ない自己研鑽に努め自己の能力の向上を行うと同時にそれを望む者を支援することとする。

9（人工知能への倫理遵守の要請）

- 人工知能が社会の構成員またはそれに準じるものとなるためには、上に定めた人工知能学会員と同等に倫理指針を遵守できなければならない。

(1) 国際的な議論のためのAI開発ガイドライン案

- AIの便益の増進及びリスクの抑制のため研究開発において留意することが期待される事項に関するG7やOECDにおける国際的な議論のための基礎となる文書として、AI開発原則(9原則)及びその内容を解説した非規制的で非拘束的なソフトローとしての指針案

(2) AIネットワーク化が社会・経済にもたらす影響

- AIネットワーク化が社会・経済にもたらすインパクト(主に良い影響、便益)及びリスクに関し、AIシステムの具体的な利活用の場面(ユースケース)を想定した評価(シナリオ分析、将来展望)を実施
- (先行的評価(災害対応、移動(車両)、教育・人材育成等10領域))及び分野別評価(公共:まちづくり、個人:健康、産業:モノの3分野))

(3) 今後の課題

- 利用者が留意することが期待される事項たる「AI利活用原則(仮称)」及びその解説からなる「AI利活用ガイドライン(仮称)」の策定に関する検討
- AIネットワーク化が社会・経済にもたらす影響に関するシナリオ分析の継続的な実施及び国際的な共有
- 人間とAIとの役割分担やステークホルダ間の役割分担、AIシステムのリスクが顕在化した場合のステークホルダ間の責任の分配の在り方など人間とAIとの関係の在り方、ステークホルダ間の関係の在り方に関する検討

基本理念

- 1. 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現すること。
- 2. AIの研究開発と利活用が今後急速に発展し、ネットワーク化されたAIシステムが国境を越えて人間及び社会に広範かつ多大な影響を及ぼすものと見込まれることから、AIシステムの研究開発の在り方について、非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダ間で国際的に共有すること。
- 3. イノベーティブでオープンな研究開発と公正な競争を通じ、AIネットワークの便益を増進するとともに、学問の自由や表現の自由といった民主主義社会の価値を最大限尊重しつつ、AIネットワークにより権利利益が侵害されるリスクを抑制するため、便益とリスクの適正なバランスを確保すること。
- 4. AI関連の技術が引き続き急速に進展していくことが見込まれる中、技術的中立性を確保する観点から特定の技術や手法に基づくAIの研究開発を阻害しないよう配慮するとともに、開発者にとって過度の負担とならないものとするよう留意すること。
- 5. AI関連技術やAIの利活用が今後とも飛躍的に発展することが期待されることから、AIネットワーク化の進展等を踏まえ、国際的な議論を通じて、本ガイドラインを不断に見直し、必要に応じて柔軟に改定すること。また、本ガイドラインの見直しに際しては関係するステークホルダの参画を得るなど、広範で柔軟な議論に努めること。

①連携の原則

- 開発者は、AIシステムの相互接続性と相互運用性に留意する。
- (主にAIシステムのリスクの抑制に関する原則)

②透明性の原則

- 開発者は、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する。

③制御可能性の原則

- 開発者は、AIシステムの制御可能性に留意する。

④安全の原則

- 開発者は、AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないように配慮する。

⑤セキュリティの原則

- 開発者は、AIシステムのセキュリティに留意する。

⑥プライバシーの原則

- 開発者は、AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する。

⑦倫理の原則

- 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。
- (主に利用者等の受容性の向上に関する原則)

⑧利用者支援の原則

- 開発者は、AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する。

⑨アカウントビリティの原則

- 開発者は、利用者を含むステークホルダに対しアカウントビリティを果たすよう努める。

(1) AIネットワーク化が社会・経済にもたらす影響

- AIネットワーク化が社会・経済にもたらすインパクト(主に良い影響、便益)
- リスクに関するシナリオ分析(行政、移動、居住の3分野)
- AIネットワーク化の進展に伴い形成されるエコシステムの展望(利用シーンの展望、ケーススタディ)
- AIネットワーク化の進展に関する考察(AIネットワーク化の観点からの課題の把握)

(2) AI利活用原則案

- AIの利活用の促進のために利活用において留意することが期待される事項に関し、非規制的・非拘束的なもの(いわゆるソフトロー)、かつ、国際的な議論のためのものとしてAI利活用原則案(10原則)

(3) 今後の課題

- 「AI開発ガイドライン(仮称)」の策定に関する検討及び国際的な議論のフォローアップ
- AIの利活用に関する指針の策定(報告書2018を踏まえた最終的なアウトプットの取りまとめ)
- AIネットワーク化が社会・経済にもたらす影響に関するシナリオ分析の継続的な実施及び国際的な共有

①適正利用の原則[安全][役割分担]

- 利用者は、人間とAIシステムとの間及び利用者間における適切な役割分担のもと、適正な範囲及び方法でAIシステム又はAIサービスを利用するよう努める。

②適正学習の原則[データ][正当性・公平性]

- 利用者及びデータ提供者は、AIシステムの学習等に用いるデータの質に留意する。

③連携の原則[連携]

- AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービス相互間の連携に留意する。また、利用者は、AIシステムがネットワーク化することによってリスクが惹起・増幅される可能性があることに留意する。

④安全の原則[安全]

- 利用者は、AIシステム又はAIサービスの利活用により、アクチュエータ等を通じて、利用者等及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。

⑤セキュリティの原則[セキュリティ]

- 利用者及びデータ提供者は、AIシステム又はAIサービスのセキュリティに留意する。

⑥プライバシーの原則[プライバシー]

- 利用者及びデータ提供者は、AIシステム又はAIサービスの利活用において、他者又は自己のプライバシーが侵害されないよう配慮する。

⑦尊厳・自律の原則[正当性・公平性]

- 利用者は、AIシステム又はAIサービスの利活用において、人間の尊厳と個人の自律を尊重する。

⑧公平性の原則[正当性・公平性]

- AIサービスプロバイダ、ビジネス利用者及びデータ提供者は、AIシステム又はAIサービスの判断によって個人が不当に差別されないよう配慮する。

⑨透明性の原則[ブラックボックス化]

- AIサービスプロバイダ及びビジネス利用者は、AIシステム又はAIサービスの入出力の検証可能性及び判断結果の説明可能性に留意する。

⑩アカウントビリティの原則[受容性]

- AIサービスプロバイダ及びビジネス利用者は、消費者的利用者及び間接利用者を含むステークホルダに対しアカウントビリティを果たすよう努める

研究課題

- 1) **研究目標**: 研究の目標となる人工知能は、無秩序な知能ではなく、有益な知能とすべきである。
- 2) **研究資金**: コンピュータサイエンスだけでなく、経済、法律、倫理、および社会学における困難な問題を孕む有益な人工知能研究にも投資すべきである。そこにおける課題として、以下のようなものがある。
 - 将来の人工知能システムに高度なロバスト性をもたせることで、不具合を起こしたりハッキングされたりせずに、私たちの望むことを行えるようにする方法。
 - 人的資源および人々の目的を維持しながら、様々な自動化によって私たちをより繁栄させるための方法。
 - 人工知能に関わるリスクを公平に管理する法制度を、その技術進展に遅れることなく効果的に更新する方法。
 - 人工知能自身が持つべき価値観や、人工知能が占めるべき法的および倫理的な地位についての研究。
- 3) **科学と政策の連携**: 人工知能研究者と政策立案者の間では、建設的かつ健全な交流がなされるべきである。
- 4) **研究文化**: 人工知能の研究者と開発者の間では、協力、信頼、透明性の文化を育むべきである。
- 5) **競争の回避**: 安全基準が軽視されないように、人工知能システムを開発するチーム同士は積極的に協力するべきである。

倫理と価値

- 6) **安全性**: 人工知能システムは、運用寿命を通じて安全かつロバストであるべきで、適用可能かつ現実的な範囲で検証されるべきである。
- 7) **障害の透明性**: 人工知能システムが何らかの被害を生じさせた場合に、その理由を確認できるべきである。
- 8) **司法の透明性**: 司法の場においては、意思決定における自律システムのいかなる関与についても、権限を持つ人間によって監査を可能とする十分な説明を提供すべきである。
- 9) **責任**: 高度な人工知能システムの設計者および構築者は、その利用、悪用、結果がもたらす道徳的影響に責任を負いかつ、そうした影響の形成に関わるステークホルダーである。
- 10) **価値観の調和**: 高度な自律的人工知能システムは、その目的と振る舞いが確実に人間の価値観と調和するよう設計されるべきである。
- 11) **人間の価値観**: 人工知能システムは、人間の尊厳、権利、自由、そして文化的多様性に適合するように設計され、運用されるべきである。
- 12) **個人のプライバシー**: 人々は、人工知能システムが個人のデータ分析し利用して生み出したデータに対し、自らアクセスし、管理し、制御する権利を持つべきである。

倫理と価値

- 13) **自由とプライバシー**: 個人のデータに対する人工知能の適用を通じて、個人が本来持つまたは持つはずの自由を不合理に侵害してはならない。
- 14) **利益の共有**: 人工知能技術は、できる限り多くの人々に利益をもたらし、また力を与えるべきである。
- 15) **繁栄の共有**: 人工知能によって作り出される経済的繁栄は、広く共有され、人類すべての利益となるべきである。
- 16) **人間による制御**: 人間が実現しようとする目的の達成を人工知能システムに任せようとする場合、その方法と、それ以前に判断を委ねるか否かについての判断を人間が行うべきである。
- 17) **非破壊**: 高度な人工知能システムがもたらす制御の力は、既存の健全な社会の基盤となっている社会的および市民的プロセスを尊重した形での改善に資するべきであり、既存のプロセスを覆すものであってはならない。
- 18) **人工知能軍拡競争**: 自律型致死兵器の軍拡競争は避けるべきである。

長期的な課題

- 19) **能力に対する警戒**: コンセンサスが存在しない以上、将来の人工知能が持ちうる能力の上限について強い仮定をおくことは避けるべきである。
- 20) **重要性**: 高度な人工知能は、地球上の生命の歴史に重大な変化をもたらす可能性があるため、相応の配慮や資源によって計画され、管理されるべきである。
- 21) **リスク**: 人工知能システムによって人類を壊滅もしくは絶滅させるリスクに対しては、夫々の影響の程度に応じたリスク緩和の努力を計画的に行う必要がある。
- 22) **再帰的に自己改善する人工知能**: 再帰的に自己改善もしくは自己複製を行える人工知能システムは、進歩や増殖が急進しうるため、安全管理を厳格化すべきである。
- 23) **公益**: 広く共有される倫理的理想のため、および、特定の組織ではなく全人類の利益のために超知能は開発されるべきである。

OECDの人工知能に関する理事会勧告

OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449. (OECD, Recommendation of the Council on Artificial Intelligence, Adopted on 22/05/2019, C(2019)34, C/MIN(2019)3/FINAL, C/M(2019)10.
<https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (2019年5月22日採択)

人権と民主的価値を尊重し、革新的で信頼できる人工知能(AI)を推進することを目的として策定されたもの

OECD加盟36カ国に加えて、アルゼンチン、コスタリカ、コロンビア、ブラジルを含む国によって採択

1)経済成長、2)貿易自由化、3)途上国支援(OECDの三大目的)に貢献することを目的とするOECDの役割と国際社会における影響力に鑑み、従来の個別の団体や研究組織が原則等の策定に向けた取り組みを実施してきたのとは異なり、今後の国際的な取り組みに大きな影響を及ぼすことが期待される

AI原則に係る理事会勧告では、信頼できるAIの責任ある管理のため、5つの補完的な価値を基本とする原則を示している。その内容は、AIが急速に進化している現状に対応して実用的かつ柔軟な基準が設定されている

OECDでは、情報の取り扱いについて、プライバシーやデジタルセキュリティリスク管理など各分野において理事会勧告が採択されているが、当該AI原則はそれらの既存のOECDの勧告等を補完するものとして位置づけられる。その適用範囲も、OECD加盟国に限定したものではなく、世界中のAI開発に適用することを念頭に置いたものである

これらの価値に基づく原則に基づき、OECDは各国政府に対して5つの「政府に求められる対応」も示している

① 包括的な成長、持続可能な開発及び幸福の促進

- AIは、あまねく広がる成長、持続可能な開発及び幸福を促進することにより、人々と地球に利益をもたらすものでなければならない。

② 人間中心の価値と公平

- AIシステムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計し、適切で安全な手段(たとえば、必要に応じて人間の介入を可能にする)を実装して、公正かつ公正な社会を確保するものでなければならない。

③ 透明性と説明責任

- AIシステムの透明性を確保し責任ある情法開示を行うことにより、AIに基づく結果を人々が理解し、それらに対抗できるようにしなければならない。

④ 堅牢性、セキュリティ及び安全性

- AIシステムは、ライフサイクル全般を通し堅牢で安全かつ安心に機能する必要がある、潜在的なリスクを継続的に評価及び管理しなければならない。

⑤ アカウンタビリティ

- AIシステムを開発、展開又は運用する組織及び個人は、上記の原則に沿って適切に動作させる責任を負わなければならない。

①AIの研究開発への投資

- 研究開発への公的及び民間投資を推進し、信頼できるAIのイノベーションを促進すること。

②AIデジタルエコシステムの育成

- デジタル社会基盤と技術、データと知識を共有する仕組みにより、利用可能なAIエコシステムを育成すること。

③AIを実現する政策環境の形成


- 信頼できるAIシステム展開への道筋を開く政策環境を確保すること。

④人的能力の構築と労働市場の変革の準備

- AIを利用する技能を人々に身につけさせ、労働者が公正に転職できるよう支援すること。

⑤信頼できるAIのための国際協力

- 国境や分野を越えて協力することで、信頼できるAIの責任ある管理を進展させること。



第3次AIブームの終焉（収束）と
ロボット法学のゆくえ

AIをめぐる法的議論 における「AI」とは

- AIとITの違いを認識した上で法的観点からの議論をしているか？
- AI + ネットワーク ≒ IoT？
- ロボット + AI + IoT = ロボット法

AIへの過大な期待

- AIブームの終焉と幻滅期の到来
- 2045年のシンギュラリティ到達論は遠すぎる将来の抽象論

AIブームにおける法律論 ≒ 情報法分野の延長線 上における 単なる「AI懸念論」

- 還元論的AIブーム法律論とロボット法学の相違

AIと自律

- AIブームに乗じて、自律(自立)できない法律論のままAIブームは終焉する

自律とは

- 法学における自律
- AI研究開発における自律

法学研究としての問題の本質の省察の必要性

- ロボット + AI + IoT
- バーチャルな問題がリアルな問題へ
- パラダイム・シフトの意義
- 無体物である情報を取り巻く問題を扱う情報法
- 有体物である自律性を有するロボットに係る新たな問題を扱うロボット法

有体物であるロボットをめぐる法律論 ≠ 製造物責任

AIブームの終焉は本来のロボット法をめぐる議論への回帰となるか



AI研究敗北論？

- AI研究では既に日本は周回遅れで負け組であるという意見が多くなりつつある
 - AIの研究開発？
 - AIをめぐる法・倫理的課題の研究？
- 研究者もAI研究開発では勝ち目がないという諦めモードに移行しつつある
- 国の政策として、AI研究敗北論を一蹴できるような科学技術政策を打ち出しているか？

AI研究敗北論的現状を打開するための提言

- (1) 科学技術の進歩に対応した制度設計・マネジメントを行う仕組み(3項目)
- (2) AIの進化と社会制度への影響も踏まえた科学技術政策の立案の必要性
- (3) データの自由な流通と適切な保護を前提とした取り組み

①科学技術の進歩に対応した制度設計を見据えた検討

- 多種多様な研究開発・イノベーションを促進するための制度設計
- ルールなきAI研究開発及び利用における弊害
 - EUは、「ルール」としてのAIの倫理指針策定に向けた具体的提案
 - 日本は、「非拘束的」な項目を例示するにとどまる
 - 日本が研究開発における原則策定について、この分野では先駆けて取り組みを行ってきたにもかかわらず、研究開発や研究者への浸透がなされていないのはなぜか
- 将来的な課題やリスクへの備え(リスクとクライシスの混同)
- 技術、機能、法的、倫理的、社会的課題(ELSI)に対応した制度設計は可能か

研究開発促進のための制度的支援

研究開発の多様性を確保する一方で、方針やルールは多様化するとわからないため、方針や政策(戦略)の**統一化**が必須

画一化しない多様かつ柔軟な議論や研究促進のため、多元的かつ多面的な検討の必要性(マルチステークホルダー)

研究組織や研究者、業界などによる分野横断的な研究の促進

バイ・デザインの対応は可能か？

行政主導による検討における課題

- 検討の重複や競合の回避
- 継続的な検討体制整備することが不可欠
 - 担当者が代わってしまうと続かなくなるというようなことがないよう産学民官の関係ステークホルダーの参画を得る必要性

複数の原則や倫理指針等が林立することによる問題

- 原則は、「原則一つ」で統一的な見解を示すべき
- 林立・濫立による原則の価値や意義の希釈化

日本は法規制が厳しいので、自動運転の車の研究開発が阻害されているという指摘を受けることが多いが本当か？

- 道交法の安全運転義務とは

「和を以て貴しと為す」法令遵守意識

研究開発に対する規制が存在しないにもかかわらず萎縮効果が生じている問題

- 研究開発と法規制
 - ヒトに関するクローン技術等の規制に関する法律
 - クローン技術又は特定融合・集合技術により作成される胚を人又は動物の胎内に移植することを禁止
 - 外為法の輸出規制
 - 基礎科学分野の研究活動において技術を提供する取引は、外国為替及び外国貿易管理法に基づく貿易関係貿易外取引に関する省令第9条第2項第10号（基礎科学分野の研究活動において技術を提供する取引）により輸出管理の対象から除外されている
- 内向的ガラパゴス化と発展的ガラパゴス化

AIの活用においては**想定外の状況でも動作を続ける必要があるため、様々な選択肢の価値評価が適切に実施される必要がある**

人間の価値観とAIの価値観の乖離によるリスクの発生や暴走の危険性だけがリスクではない

AIの進化により人間と同様の価値観を持たせることで生ずる社会的にも予期しない不利益が生ずるなど多大な損失や被害が発生する可能性

予見可能性とは

- 予見していないがゆえにどのような問題が発生するか未知である一方で、当該情報技術を用いることで影響を受ける可能性があるステークホルダーは、社会全般に及ぶ可能性がある

社会制度の根幹にも関わる問題でもあるため、一部のステークホルダーのみが影響を受ける問題ではないことに留意した検討がなされるべき

バーチャルとリアルの融合 (Society 5.0、データ覇権争い) により、データの自由な流通と保護が一層重要な意味を持つ

- 経済政策や安全保障政策との関係で、データの囲い込み、いわゆるデータローカライゼーションを目的とした制度の整備が国際的になされる傾向にある
- 日本の政策においては、データの自由な流通と保護を基調とする原則に基づく持続可能な発展的な取り組みを継続することが不可欠

AI原則は機能するか？ — 自律型致死兵器システム(LAWS)に関する法的・倫理的側面からの基本原則提案を手がかりに —

- AIや自律型ロボットの研究開発の進展に呼応して、エマージング・テクノロジーの新たな可能性とそれに伴う問題に対応すべく、国内外で様々なルール作りの試みや原則等の提案がなされている。原則策定に向けた取り組みにおいて日本は先行してきたが、AIの研究開発及び利用における汎用的な原則として機能させることはできるであろうか。AIの将来的な運用において最も深刻な懸念が生ずる可能性が高い「自律型致死兵器システム(LAWS)」をめぐる法的・倫理的側面からの基本原則提案を手がかりに、AI原則策定への取り組みを省察したい。

① 機微技術管理への取り組みの必要性

- 技術優位性の毀損や技術の脆弱性が安全保障上の重大な懸念事項となりつつある
- 国際的な動向として、安全保障を理由とする機微技術の流出防止策や自国産業を中心に据えた産業政策が拡大
- 軍事転用可能な技術の拡散防止の観点から、機微技術の流出防止策を講じることは国際的な義務
- ネットワーク上を流通するパーソナルデータも含む安全保障に関わり得るデータの保護についても、機微技術管理に包含される論点として検討が必要であるとともに、経済成長や技術革新を阻害することを避ける必要がある
- 実効性のある機微技術管理のあり方について検討すべき

② 自律型兵器の国際的規制に向けた取り組み

- AIや自律型ロボットをはじめとするエマージング・テクノロジーの急速な発展と産業における利用が議論される一方で、自律型致死兵器システム(LAWS)の研究開発も進展している
- 国際的な議論としては、特定通常兵器使用禁止制限条約(CCW)の枠組みにおける規制について検討がなされ原則指針の提案にまで至っている
- 本報告では、自律型致死兵器システムについて、LAWSに適用される国際人道法、国連軍縮研究所(UNIDIR)の報告書、CCWのLAWS非公式専門家会議から政府専門家会合(GGE)における議論の状況、提案された基本原則の内容、法的・倫理的側面をめぐる検討状況を踏まえて、LAWSの国際的な規制に向けた取り組みの現状について報告する

新興技術に関する新たな可能性とそれに伴う問題は国内外で様々な側面から検討されている

日本国内における議論は？

- AIやロボットの利用と産業振興の観点からの検討に関するものが一般的
 - 例えば、ドローンの技術開発に関するもの
 - 多種多様なサービスやビジネスでそれらを活用する提案
 - 宅配サービスにおける利用への期待
 - インフラ管理における点検監視
 - 測量や建設現場・農業等における利用
 - 個人の空撮などの趣味に至るまで

国外における議論は？

- 無人戦闘航空機など軍事利用も含む研究開発
 - 軍事研究や宇宙開発分野の研究予算の額が突出して多いことから当然
 - 国内においても、新たな通常兵器であるドローンやロボット兵器といった人工知能搭載兵器をどのように位置づけていくべきかを検討する岩本誠吾「ドローンと法規制」防衛技術ジャーナル35(7)、PP.16-19(2015)をはじめとして、「先端技術の開発・規制のための適切な国際ルールの提案」「先端技術に関する国際政治経済の実態把握」「先端技術情報データベース構築」を三本柱に新興技術の利用をめぐる問題を研究し、当該研究の関連研究成果を示している、岩本誠吾「革新的技術の国際法政治経済への影響の分析」京都産業大学総合学術研究所報13、PP.137-141(2018)がある。
- ドローン(Drone)という用語の出自
 - 「人間が直接操作することなく自律的に飛行することができる無人航空機」を意味する
 - John Villasenor, *What Is a Drone, Anyway?* <<https://blogs.scientificamerican.com/guest-blog/what-is-a-drone-anyway/>>Scientific American, on April 12, 2012.
 - 無人航空機一般はUAV(Unmanned Aerial Vehicle)と表記される
 - ドローンはUAVに含まれるものの、自律的に飛行できない無人航空機も含むことからUAVのほうが概念上はドローンよりも広義であると解する指摘もある
 - 「有人による地上からの遠隔操作」も含む無人航空機一般はUAS (Unmanned Aircraft Systems)
 - Global Hawkのように高度な操縦技能が必要なものをRPA(Remotely Piloted Aircraft)と表記することもある
 - Shawn Herrick, *What's The Difference Between A Drone, UAV and UAS* <<https://botlink.com/blog/whats-the-difference-between-a-drone-uav-and-uas>>November 14, 2017.
- LAWSの研究開発の現状については、富川英生「各国の軍事イノベーションと自律システムの開発」『第3相殺戦略』を見据えた新興技術の研究開発動向」日本安全保障貿易学会第27回研究大会
- LAWSに至る兵器開発の歴史を紹介した上で、定義や自律性とは何かを考察するものとして、Nicholas W. Mull, *The Robotization of Warfare With Lethal Autonomous Weapon Systems (Laws): Mandate of Humanity or Threat to It?*, 40 Hous. J. Int'l L. 461 (2018).

3. 安全保障関係研究に対する日本の学术界の反応

安全保障研究に対する学术界の姿勢は？

- 競争的資金制度として2015年から防衛装備庁が安全保障技術研究推進制度を実施している
- 「軍事的安全保障研究に関する声明 - 日本学術会議」(平成29年(2017年)3月24日)の公表以降、安全保障関係の研究が積極的に推進されている状況とは言えない

声明公表後の報道例

- 軍事応用研究採択14件(日本経済新聞2017年8月30日夕刊18ページ)
- 軍事研究しない、京大が方針決定 HPで明文化(朝日新聞2018年3月29日夕刊2面)
- 軍事研究指針整備 3割のみ(東京新聞2018年4月4日朝刊3面)
- 軍事研究規制の道筋(毎日新聞2018年6月4日朝刊4面)
- 北大 防衛省研究を辞退(読売新聞2018年6月9日朝刊33面)

その他参考情報

- 平成30年度安全保障技術研究推進制度公募要領 http://www.mod.go.jp/atla/funding/koubo/h30/h30koubo_full.pdf
- 平成29年度の安全保障技術研究推進制度 採択情報 <http://www.mod.go.jp/atla/funding/kadai/h29kadai291222.pdf>
 - 大規模研究課題(タイプS)6件
 - 小規模研究課題(タイプA・B)8件

自律型兵器を用いた安全保障に関わる制度的な課題研究の必要性


- 文字通りの軍事技術開発研究に研究者が従事することに反対の意思を表明すると同時に、自律型兵器を用いた安全保障に関わる制度的な課題についてまでも、新たな軍事技術を利用した軍事研究開発をめぐる議論と同一視し研究への従事に萎縮することは避けなければならない

AI研究開発における規制のあり方

- 自律型兵器の規制をめぐる議論は、安全保障分野における問題にとどまらず、AI研究開発における規制のあり方や、その検討過程における日本の位置づけを認識する機会としても重要

原則策定に向けた取り組み

- AIや自律型ロボットの日本における研究開発が後塵を拝しつつあると指摘される機会が増える一方で、そのルール作りでは先行しアドバンテージを維持している現状において、AIや自律型ロボットの開発や利用に伴う原則策定に向けた取り組みなどのノウハウを生かすことができる絶好の機会
- 安全保障分野や軍縮問題で日本がイニシアチブを取ることができるのかという点については、諸外国との安全保障環境における我が国の位置づけからすると懐疑的にならざるを得ない
- 一方で、そのような原則策定については日本国内における検討を積極的に活用するとともに、国際的な枠組みにおける議論への参画も可能



① 機微技術管理への取り組み

1. 国際秩序の変容と我が国がとるべき経済政策へのインプリケーション


1. 1. 国際秩序の変容

- 国内格差の拡大や、異なる政治・経済体制の台頭に対する不満や不安から、自国第一主義の傾向が顕在化
- WTO 体制、EU 統合といった「民主主義」や「自由で公正な競争」といった価値観を軸とする国際秩序が、国際政治、通商、テクノロジーなどあらゆる分野で揺らいでいる
- 我が国としても、引き続き WTO を中心とする国際協調主義を基調としつつも、「安全保障と一体となった経済政策」が必要となっている
- 価値観・理念を共有する有志国との連携を進めながら、我が国の経済的優位性をどのように維持・向上させていくかが問われている

1. 2. 「安全保障と一体となった経済政策」の方向性

(1) 統合的アプローチの重要性

- 世界的に、技術優位性の毀損や技術の脆弱性が安全保障上の懸念であるとの位置づけの下、安全保障を理由とする機微技術の流出防止策や自国産業を中心に据えた産業政策が拡大
- 軍事転用可能な技術の拡散防止の観点から、政府として然るべき機微技術の流出防止策を講じることは我が国が果たすべき国際的な義務
- グローバリゼーションやオープン・イノベーションの結果として生じている流出経路の多様化に対応しつつも、経済成長や技術革新を阻害することを避ける必要
- 技術優位性の確保や技術の脆弱性の解消は、機微技術の流出防止策を手当するだけでは実現できない
 - 「機微技術」＝「我が国の安全保障を確保し、もって経済の健全な発展を実現する上で、その優位性を保ち、脆弱性を解消すべき重要な技術(critical technology)」と定義



② 自律型兵器の国際的規制に向けた取り組み

1. 自律型致死兵器システムとは（用語について）

LAWS

• Lethal Autonomous Weapon Systems

- LAWSという用語で当該兵器に関する規制をめぐる検討状況を紹介した先行研究として、佐藤丙午「致死性自律兵器システム(LAWS)をめぐる諸問題（ロボット技術と戦争）」国際安全保障42(2)、PP.1-14(2014)。
- Rebecca Crootof, Autonomous Weapon Systems and the Limits of Analogy, 51 Harv. Nat'l Sec. J. 51,57 (2018)では、①完全自律、②準自律、③自動、④非自動(inert)にレベル分け。Rebecca Crootof, The Killer Robots Are Here: Legal and Policy Implications, 36 Cardozo L. Rev. 1837, 1842(2015)においても、定義、定義の明確化の必要性、自律のレベル分けが示されている。
- Matthew Ormsbee, Centaur Warriors: A Legal Analysis of Autonomous Systems in Military Operations, 54 Willamette L. Rev. 47, 50 (2017)は、米国防総省の認識として、①準自律システム(SAS)、②人的監視下にある自律システム(HSAS)、③自律兵器システム(AWS)に分類。
- 定義づけの試みとしては、Michael N. Schmitt, Autonomous Weapon Systems and International Humanitarian Law: A Reply to the Critics, Harv. Nat'l Sec. J. Features 1, 4 (2013)では、LAWS are weapon systems that are capable - at a base minimum - of detecting, identifying, selecting, and lethally engaging human targets in war without direct human supervision or controlとしている。

AWS

• Autonomous Weapons Systems

LAWs

• Lethal Autonomous Weapons

LAR

• Lethal Autonomous Robots

- LARという用語で当該兵器に関する規制をめぐる検討状況を紹介した先行研究として、岩本誠吾「致死性自律型ロボット(LARs)の国際法規制をめぐる新動向」産大法学47(3/4)、PP.330-363(2014)。

Robotic Weapons

Killer Robots

現時点において研究開発に従事している国々
アメリカ、イギリス、イスラエル、韓国、中国、ロシアなど

Francis Grimal, Combat Drones: Hives, Swarms, and Autonomous Action?, J Conflict Security Law (2018) 23 (1): 105では、ドローンを軍事目的で利用している国の現状を紹介

LAWSの禁止に賛同する国は2018年11月22日時点で29カ国あるとされる。

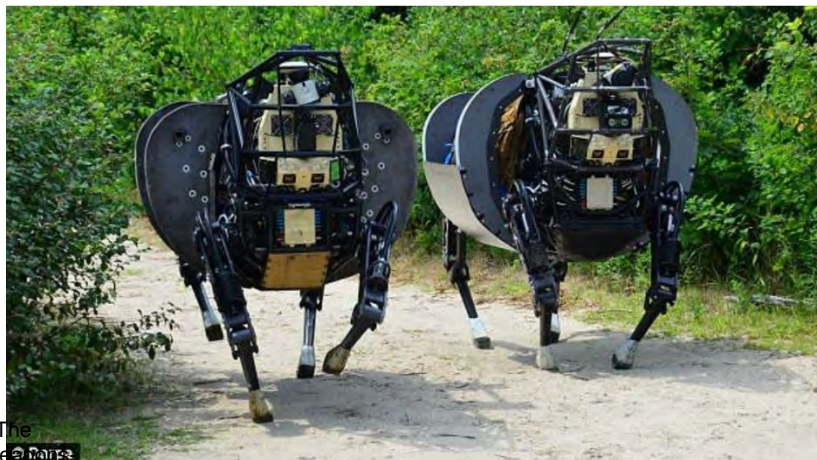
禁止に賛同する国々のリスト<https://www.stopkillerrobots.org/wp-content/uploads/2018/11/KRC_CountryViews22Nov2018.pdf>

1-1 LAWSの具体例と仮想例

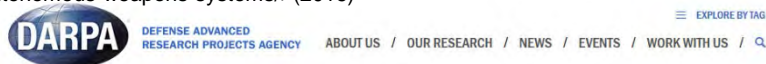


The Defender, a U.S. robotic platform that performs reconnaissance, surveillance, targeting, and threat neutralization tasks. Image: US Air Force

Fergus Kelly, Killer robots? UN to hold talks on Lethal Autonomous Weapons Systems, The Defense Post<<https://thedefensepost.com/2017/11/10/killer-robots-lethal-autonomous-weapons-systems-un/>>(2017)



Tasneem Jamal, Workshop: Meaningful Human Control and Lethal Autonomous Weapons Systems<<http://ploughshares.ca/2016/06/workshop-meaningful-human-control-and-lethal-autonomous-weapons-systems/>>(2016)



Defense Advanced Research Projects Agency > Program Information

Gremlins

Mr. Scott Wierzbanski

Gremlins

For decades, U.S. military air operations have relied on increasingly capable multi-function manned aircraft to execute critical combat and non-combat missions. Adversaries' abilities to detect and engage those aircraft from longer ranges have improved over time as well, however, driving up the costs for vehicle design, operation and replacement. An ability to send large numbers of small unmanned air systems (UASs) with coordinated, distributed capabilities could provide U.S. forces with improved operational flexibility at much lower cost than is possible with today's expensive, all-in-one platforms—especially if those unmanned systems could be retrieved for reuse while airborne. So far, however, the technology to project volleys of low-cost, reusable systems over great distances and retrieve them in mid-air has remained out of reach.

To help make that technology a reality, DARPA has launched the Gremlins program. Named for the imaginary, mischievous imps that became the good luck charms of many British pilots during World War II, the program envisions launching groups of UASs from existing large aircraft such as bombers or transport aircraft—as well as from fighters and other small, fixed-wing platforms—while those planes are out of range of adversary defenses. When the gremlins complete their mission, a C-130 transport aircraft would retrieve them in the air and carry them home, where ground crews would prepare them for their next use within 24 hours.

The gremlins' expected lifetime of about 20 uses could provide significant cost advantages over expendable systems by reducing payload and airframe costs and by having lower mission and maintenance costs than conventional platforms, which are designed to operate for decades.

The Gremlins program plans to explore numerous technical areas, including:

- Launch and recovery techniques, equipment and aircraft integration concepts
- Low-cost, limited-life airframe designs
- High-fidelity analysis, precision digital flight control, relative navigation and station keeping

The program aims to conduct a compelling proof-of-concept flight demonstration that could employ intelligence, surveillance, and reconnaissance (ISR) and other modular, non-kinetic payloads in a robust, responsive, and affordable manner.



From 'I Robot' trailer/via YouTube

Artificial intelligence experts urge UN to impose curbs on autonomous weapons<<https://scroll.in/latest/847989/artificial-intelligence-experts-urge-un-to-impose-curbs-on-autonomous-weapons>>

1-2 自律兵器と自動兵器防衛システムとの違い

米海軍: MK15 Phalanx ファランクス接近防空システム (Close in Weapons System, CIWS)

米: 対ロケット・野戦砲・迫撃砲システム (The CounterRocket, Artillery, and Mortar System, C-RAM)

イスラエルのIron Dome: ガザ境界線などに実戦配備

CIWS

• <https://dod.defense.gov/News/Special-Reports/Remembering-September-11th-2017/Videos/videoid/558866/dvpTag/CIWS/>

CRAM

• <https://www.youtube.com/watch?v=9HR9BkL5S6k>

Iron Dome

• <https://www.youtube.com/watch?v=BsrEY3YZqsl>

IAI Mini Harpy

• <https://www.youtube.com/watch?v=cBmhJ-jrzrc>

特定通常兵器使用禁止制限条約

(Convention on Certain Conventional Weapons : CCW)

Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons which may be deemed to be Excessively Injurious or to have Indiscriminate Effects (with Protocols I, II and III) Geneva, 10 October 1980.

国際的な武力紛争の犠牲者の保護を定めるジュネーブ諸条約 第一追加議定書(追加議定書)

- 国際法の適用関係に関する解説は、福井康人「新たな技術と国際法の適用可能性：自律型致死性兵器システム(LAWS)を事例として(特集 21世紀における戦争と平和)」世界法年報(36)、PP.154-185(2017)
- Lieutenant Colonel Christopher M. Ford, *Autonomous Weapons And International Law*, 69 S.C. L. Rev. 413(2017)も同様の視点
- LAWS規制については、川口礼人「今後の軍事科学技術の進展と軍備管理等に係る一考察：自律型致死兵器システム(LAWS)の規制等について」防衛研究所紀要19号1巻、PP.213-231(2016)。岩本誠吾「ロボット兵器と国際法」弥永真生、宍戸常寿(編)『ロボット・AIと法』有斐閣(2018)、齋藤洋「自律型致死性兵器システム(LAWS)の責任問題に関する現行国際法上の規制と課題」憲法研究50号、PP.67-86(2018)。
- Alan L. Schuller, *At the Crossroads of Control: The Intersection of Artificial Intelligence in Autonomous Weapon Systems with International Humanitarian Law*, 8 Harv. Nat'l Sec. J. 379 (2017)では、国際人道法に基づくLAWS規制について論じた上でその限界を指摘。

第36条「新たな兵器」

- 締約国は、新たな兵器又は戦闘の手段若しくは方法の研究、開発、取得又は採用に当たり、その使用がこの議定書又は当該締約国に適用される他の国際法の諸規則により一定の場合又はすべての場合に禁止されているか否かを決定する義務を負う。
- 合法性について、兵器自体、その使用方法の2つの側面から定めている
- 研究、開発、取得、採用の4段階で国際人道法に適合しているか否かを決定する義務を締約国に課している

追加議定書の主な3つの原則

- (1) 戦闘方法・手段は無制限ではなく、過度の傷害や無用の苦痛を与えることへの禁止原則（第35条）
- (2) 文民と戦闘員とを、民用物と軍事目標とを常に区別し、軍事目標だけに限定して攻撃する区別原則（第48条、第51条、第52条等）
- (4) 予測される文民の巻き添えの損害と予期される軍事的利益との均衡原則（第51条等）
 - 交戦規定（ROE: rules of engagement）

特定通常兵器使用禁止制限条約(Convention on Certain Conventional Weapons: CCW)として1980年にジュネーブにて採択。1983年に発効。

条約の構成

•本条約は、手続等基本的事項につき規定した本体条約(締約国125)及び個別の通常兵器について規制する以下の5つの附属議定書から構成される

1) 検出不可能な破片を利用する兵器に関する議定書(議定書I; 1983年発効)

•検出不可能な破片によって傷害を与えることを第一義的な効果とする兵器の使用を禁止している(使用の全面禁止)。

2) 地雷, ブービートラップ等の使用の禁止又は制限に関する議定書(議定書II; 1983年発効, 1996年に改正(改正議定書II), 1998年発効)

- 1983年の議定書は, 対人地雷が主に使用される内乱には適用されず, また, 探知不可能な地雷等を禁止していない等の問題点を内包していたが, 1996年に改正された議定書は内乱にも適用され, 一定の地雷(探知不可能なもの又は自己破壊機能を有さないもの)の使用制限や移譲の規制が盛り込まれるなど規制が強化された。
- その後, 本件改正議定書に基づく部分的な禁止では対人地雷問題の抜本的な解決には至らないとするNGO等によって, CCWの枠外でオタワ・プロセスが開始され, 対人地雷全面禁止条約が作成された(1997年署名, 1999年発効)。

3) 焼夷兵器の使用の禁止又は制限に関する議定書(議定書III; 1983年発効)

•文民及び民用物をナパーム弾等の焼夷兵器による攻撃目標とすること, 人口周密地域にある軍事目標を攻撃目標とすること等を禁止している(部分的な使用規制)。

4) 失明をもたらすレーザー兵器に関する議定書(議定書IV; 1998年発効)

•永久に失明をもたらすように特に設計されたレーザー兵器の使用及び移譲の禁止等を規定している(使用の全面禁止)。

5) 爆発性戦争残存物(ERW)に関する議定書(議定書V; 2006年発効, 我が国は未締結)

•主に不発弾等の危険を最小化するために, 紛争後の対応措置や, 不発弾の発生を最小化するための技術的予防措置を規定している。

特定通常兵器使用禁止制限条約（Convention on Certain Conventional Weapons: CCW）におけるLAWSをめぐる議論の現況について

外務省「特定通常兵器使用禁止制限条約関連会合」

<<https://www.mofa.go.jp/mofaj/gaiko/arms/ccw/index.html>>に掲載されている会議の議事概要を抜粋して議論の状況を以下確認

1. 締約国会議（2013年11月11日から15日）

- 自律型致死兵器システム（Lethal Autonomous Weapons Systems）に関し、2014年5月13日から16日まで、非公式専門家会合を開催し、その結果を2014年の締約国会合に報告することが決定された。
- 人道的配慮と安全保障の双方の観点から通常兵器の使用の規制を検討するCCWの枠組みにおいて自律型致死兵器システムに関する非公式専門家会合の開催が決定されたことは有意義との評価

1. 概要

- (1) 5月13日から16日まで、ジュネーブにて、特定通常兵器使用禁止制限条約(CCW)自律型致死兵器システム(LAWS)に関する非公式専門家会合が開催され、我が国を含む83か国の他、国際機関、NGOが参加した。我が国からは、佐野軍縮代表部大使を団長として、佐藤丙午拓殖大学国際学部教授、外務省、経済産業省、防衛省関係者が出席した。
- (2) 今次会議においては、一般討論の他、各セッション(技術、倫理、法律、軍事)において専門家による発表が行われ、それに対して出席者から質問や意見が出された。我が国からは、和久田経済産業省航空機武器宇宙産業課企画官及び佐藤丙午拓殖大学国際学部教授が、技術及び軍事のセッションにおいてプレゼンテーションを行った。
- (3) 今次会合においては、LAWSに関わる諸問題として、ロボット技術の自律性の程度、軍事的有用性、国際人道法の適用、LAWSの使用に伴う責任の所在、LAWSの使用にかかる倫理的問題等について議論された。本会合において議長の権限で作成された議長サマリーは、本年11月に開催されるCCW締約国会議において提出され、今後の議論の取り進め方について決定されることとなっている。

2. 評価

- (1) 今次会合は、LAWSの問題点等について各国が意見交換を行い、共通の理解を深めることに役立ち有意義であった。また、我が国としても、一般討論の他、全てのセッションにおいて積極的に発言を行い、さらに技術及び軍事のセッションにおいて発表を行う等の貢献を行うことができた。
- (2) 本年11月のCCWの枠組み締約国会議に提出される予定の議長サマリーにおいて、我が国の主張するロボットの民生技術の重要性が指摘されたこと、また多くの国(我が国を含む)がLAWSに関するCCWの枠組みにおける議論の継続を望んでおり、次回締約国会合の決定に従って、今後の議論を行っていくことが決定される旨明記されたことは、有意義であった。

3. 締約国会議（2014年11月10日から14日）

- 1. 概要
 - 自律型致死兵器システム(LAWS:Lethal Autonomous Weapons Systems)に関し、2015年4月13日から17日まで、第2回非公式専門家会合を開催することが決定された。また、履行支援ユニット(ISU)が作成した報告書(条約の普遍化の促進、自律型致死兵器システム、スポンサーシップ・プログラム、履行支援ユニット)が提出され、これに基づいて議論が行われた。
- 2. 評価
 - 人道的配慮と安全保障の双方の観点から通常兵器の使用の規制を検討するCCWの枠組みにおいて自律型致死兵器システム(LAWS)に関する第2回非公式専門家会合の開催が決定されたことは、有意義であった。

1 概要

- (1) 4月13日から17日まで、ジュネーブにて、特定通常兵器使用禁止制限条約(CCW)自律型致死兵器システム(LAWS)に関する第2回非公式専門家会合が開催され、我が国を含む74か国の他、国際機関、NGOが参加した。我が国からは、外務省、防衛省関係者が出席した。
- (2) 今次会議においては、一般討論の他、各セッション(技術問題、LAWSの特徴、自律性の向上による国際人道法に対するあり得べき課題、横断的な課題、透明性及び今後の取組)において研究機関等の専門家による発表が行われ、それに対して出席者から質問や意見が出された。我が方からは、一般討論においてロボット技術の汎用性は引き続き重視すべき論点であること等を内容とするステートメントを行った。
- (3) 今次会合においては、LAWSの開発を確実にする発表はなく、LAWSに関わる諸問題として、LAWSの定義の明確化、LAWS使用に伴う責任の所在、LAWSが人間の生死に関わる意思決定を行うことの倫理上の問題、LAWSへの国際人道の適用、LAWSに関する透明性措置の可否、今後の取組のあり方等について議論された。本会合において作成された議長サマリーは、本年11月に開催されるCCW締約国会議に報告されることとなった。

2 評価

- (1) 今次会合は、LAWSの問題点等について各国が意見交換を行い、第1回非公式会合(2014年5月)での議論をより深めることができ、有意義であった。今次会合での議論では、LAWSの規制又は禁止を求める国や専門家がいる一方で、LAWSがもたらす倫理上の問題等に懸念を共有しつつ、具体的な規制や禁止については慎重であるべきであり、引き続き議論すべきとの立場を表明する国もあったことが注目された。
- (2) 本年11月のCCWの枠組み締約国会議に報告される予定の議長サマリーは、引き続き議論を深めることが必要であるとの全般的な理解がある一方で、議論を継続する方途(非公式専門家会合、又は政府専門家会合の形式で実施するか等)については異なった考えがある旨を指摘している。本年11月の締約国会議では、今後の議論の取り進め方についても検討が行われると思われる。

概要

- 自律型致死兵器システム (Lethal Autonomous Weapons Systems) に関し、2016年4月11日から15日まで、非公式専門家会合を開催し、その結果を第5回運用検討会議 (2016年12月12日～16日に開催予定) に報告することが決定された。
- 履行支援ユニット (ISU) が作成した4種類の報告書 (条約の普遍化の促進、自律型致死兵器システム、スポンサーシップ・プログラム、履行支援ユニットの活動報告) が提出され、これに基づいて議論が行われた。

1 概要

- (1) 4月11日から15日まで、ジュネーブにて、特定通常兵器使用禁止制限条約(CCW)自律型致死兵器システム(LAWS)に関する第3回非公式専門家会議が開催され、我が国を含む94か国の他、国際機関、NGOが参加した。我が国からは、外務省、防衛省関係者が出席した。
- (2) 本会議においては、一般討論の他、各セッション(自律性の考察、LAWSの作業場の定義に向けて、国際人道法上の課題、人権と倫理問題、安全保障)において研究機関等の専門家による発表が行われ、それに対して出席者から質問や意見が出された。我が方からは、一般討論においてロボット技術の汎用性は、引き続き重視していくべき論点であること等を内容とするステートメントを行った。これに関し、多くの国が、LAWSに関して取られる規制措置が民生分野の開発を妨げることになってはならない旨主張した。
- (3) 本会合においては、完全自律型兵器は将来開発され得るという意見がある一方で、そのような兵器は開発されることはないとする意見もあった。他方、LAWSの国際法(特に国際人道法)の適用の重要性については、ほぼ全ての国が主張していた。また、多くの国が、武器の法的審査が国際人道法の遵守を確保するために重要であることを主張した。その他、LAWSの使用が人権、特に人間の尊厳等に対して、深刻な影響を及ぼすことを危惧する意見が多数述べられた。
- (4) 本会合において、LAWSに関する政府専門家会議の設立を含むCCW第5回検討会議(5RC)への勧告がコンセンサスで採択された。また、議長の権限で作成された議長サマリーは、本年12月に開催される5RCに報告されることとなった。

2 評価

- (1) 本会合は、LAWSの問題点等について各国が意見交換を行い、第1回及び第2回非公式会議(2014年及び2015年)での議論を踏まえ、より政策的な議論を深めることができ、有意義であった。
- (2) 本会合においてコンセンサスで採択された5RCへの勧告には、政府専門家会議(GGE)を開催することが明記されたが、2017年から開催されるGGEにおいては、LAWSの特徴についての特定及び作業上の定義の明確化、国際人道法等の国際法の関連原則の適用及び遵守について、専門的な議論を集中して実施することが期待される。第5回履行検討会議では、GGEに関する今後の具体的な議論の取り進め方についても検討が行われる予定である。

概要

- 11月13日から17日まで開催された自律型致死兵器システム(Lethal Autonomous Weapons Systems: LAWS)に関する政府専門家会合からの報告書についても検討が行われた。
- 同会合の勧告に基づき、政府専門家会合の2018年における開催が決定された。
- その他、条約の着実な履行や普遍化へむけた締約国各国の努力を呼びかける等の最終文書を全会一致で採択した。

1.

- 4月9日から13日まで、ジュネーブにおいて、特定通常兵器使用禁止制限条約(CCW)の枠組で、自律型致死兵器システム(LAWS)に関する政府専門家会合(GGE)が開催され、我が国を含む約80か国の他、国際機関、NGOが参加しました。我が国からは、高見澤将林軍縮代表部大使を団長として、佐藤丙午拓殖大学国際学部教授、外務省、防衛省関係者が出席。

2.

- 今回の政府専門家会合は、過去3年(2014年から2016年)にわたり開催された非公式専門家会合及び昨年11月の初めての政府専門家会合の成果を踏まえつつ、一般討論における各国、団体の代表者による意見表明の他、LAWSの特徴、人間と機械の相互関係さらにはLAWSに係る先端技術について議論が行われ、LAWSに関する認識の共有が図られた。

3.

- 我が国からは、一般討論等において、人間が関与しない完全自律型の致死性兵器の開発を行う意図は有していないとの従来からの考え方を示した上で、日々発展を続ける人工知能(AI)を含む先端技術は、正しい使い方をすれば人類にとって明るい未来をもたらす可能性を秘めている一方で、誤って使用された場合、人類への大きな脅威になり得る旨述べました。そのような観点から、ロボット技術やAI技術において先進的な技術を有する我が国としては、民生分野における同技術の健全な発展を妨げるものであってはならず、今次GGEの中で民間部門の意見も参考に議論することは、より現実的な認識を共有するのに有意義である旨述べている。

4.

- また、自律型致死兵器の使用にあたっては、人による有意の制御(MHC: Meaningful Human Control)が必要不可欠であるという考え方のもと、いかなる次元で人間の介入がなされるべきなのかといった議論を通じ、国際的な共通理解を得られるよう継続的な議論を支持する旨述べている。

5.

- 今次政府専門家会合は、昨年11月のCCW締約国会議の決定に基づいて開催されたものであり、8月下旬には本年第2回目の会合が予定されています。我が国は、今後とも本件分野における国際的な議論に積極的かつ建設的に参加していくとしている。

1.

- 8月27日から31日まで、ジュネーブにおいて、特定通常兵器使用禁止制限条約（CCW）の枠組の下、自律型致死兵器システム（LAWS）に関する政府専門家会合（GGE）が開催され、我が国を含む約70か国の他、国際機関、NGOが参加しました。我が国からは、高見澤将林軍縮代表部大使を団長として、外務省及び防衛省関係者他が出席。

2.

- 今回の政府専門家会合は、4月に開催された第1セッションに続く2018年会期の第2セッションであり、前回セッションと同様の議題の下、LAWSの特徴、LAWSの使用における人間の関与、国際人道法上の課題等について、活発な議論が行われた。

3.

- 会合において、我が国は、人間の関与が及ばない完全自律型の致死性兵器の開発を行う意図は有しておらず、兵器については、人間による関与が必須であるとの立場を表明。
- また、LAWSにも用いられるAI技術は、経済や社会の様々な分野で利用されており、今後も急速な発展や、多大なメリットをもたらすことが見込まれており、こうした技術の発展や技術革新を阻害することのないよう、LAWSに関しては冷静かつバランスの取れた議論を国際的に継続していく必要がある旨述べている。

4.

- 会合の最終日には、報告書が採択され、本年の2つのセッションにおける議論の概要がとりまとめられるとともに、2019年も引き続きCCWの枠組みで議論を継続していくことが提言された。この提言は、本年11月のCCW締約国会議において審議される予定。

1.

- 本5日午前11時10分頃から約1時間、外務省は、自律型致死兵器システム（LAWS）に関する有識者との第3回意見交換会を開催。

2.

- LAWSについては、特定通常兵器使用禁止制限条約の枠組みの下で昨年11月に引き続き、本年4月に第2回、8月にジュネーブにおいて第3回政府専門家会合が実施され、LAWSの特徴、人間と機械の相互関係さらにはLAWSに係わる先端技術等について活発な議論が行われてきている。
- 本件意見交換会は、こうした国際的に深まる議論も踏まえ、外務省として一の研究を進める必要があるとの考えから、本年3月、7月に引き続き開催したもの。

3.

- 今回の意見交換会では、国際人道法や倫理的側面からのLAWSの問題、今後の国際的な規制のあり方等について議論を行った。

4.

- 外務省としては、LAWSに関する更なる研究を進め、今後とも本件分野における国際的な議論に積極的かつ建設的に参画していく考え。

概要

- 本年4月及び8月に開催された自律型致死兵器システム(Lethal Autonomous Weapons Systems: LAWS)に関する政府専門家会合から締約国会議へ提出された報告書が承認された。
- 同政府専門家会合の2019年における開催が決定された。
- その他, CCW枠組条約及び各附属議定書の着実な履行や普遍化へ向けた締約国各国の努力を呼びかけることを主要な内容とする最終文書がコンセンサスで採択された。

4. 国連軍縮研究所 (UNIDIR) の2017年報告書

報告書の名称

- United Nations Institute for Disarmament Research(UNIDIR), The Weaponization of Increasingly Autonomous Technologies: Concerns, Characteristics and Definitional Approaches, No. 6, 2017.

報告書策定の経緯

- 2016年にCCW条約の締約国会議において、政府の専門家グループ(GGE)の設立を決定
- LAWSの実用的な定義の検討に着手することが目的
- LAWSをめぐる国際的な規制のあり方に関する議論をまとめたもの
- CCWにおける検討における基礎資料
- UNIDIRが公表しているAIやLAWSに関する報告書は、この問題を議論するにあたっての基礎となる情報を提供するとともに、具体的な検討事項を議論するための体系的な資料と位置づけられるもの

2018年のAIの基本に関する解説報告書

- United Nations Institute for Disarmament Research(UNIDIR), The Weaponization of Increasingly Autonomous Technologies: Artificial Intelligence, No. 8, 2018.
 - CCWにおける検討の基礎資料として2017年の報告書の後に公表されたもの
- LAWSの議論の前提となるAIそのものについて理解することを目的とした構成
 - ①人工知能とは何か？
 - ②機械学習
 - ③深層学習
 - ④コンピューティング・リソース
 - ⑤その他のAIおよび機械学習の方法
 - ⑥自律性の向上
 - ⑦特化型と汎用型AI
 - ⑧スーパーインテリジェンス

1. 国際的な議論として提起されている懸念事項の整理
2. LAWSに関連して生じている特有の問題の探求
3. 異なる定義付けのアプローチの解説
4. 定義の実例と懸念事項の特徴
5. 結論

自律兵器システムに関する国際的な議論の初期の段階は、人権と法的な問題を中心とした議論

現在は、政府が自律性を有する武器システムの開発を実際に進展させているため、当初とは異なる問題と懸念事項の深刻化が顕著になりつつある

本報告書はAIの進化に伴い自律的な動作が可能な技術が発達することに伴う具体的な懸念事項を示している

- 1899年第2次ハーグ陸戦条約の前文に記され、1977年追加議定書IIIに定められたマルテンス条項
- 「実効性を有する法律が適用されない場合であっても、人は人道の原則に基づいて保護され、公共の良心に従う。」

4-3 UNIDIR2017年報告書が示す懸念事項

| | |
|-------------------|---|
| ① 人権と倫理 | (Human rights and ethics) |
| ② 合法性 | (Legality) |
| ③ 技術 | (Technological) |
| ④ 拡散及び軍拡競争 | (Proliferation and arms-racing) |
| ⑤ 戦略的安定性 | (Strategic stability) |
| ⑥ 技術のデュアルユースの側面 | (The dual-use nature of the technologies) |
| ⑦ 運用上の懸念 | (Operational concerns) |
| ⑧ 意図しないリスクと安全性の問題 | (Unintentional risk and safety issues) |
| ⑨ 「静的」又は「動的」自律性 | (Autonomy “at rest” or “in motion”) |
| ⑩ 懸念の切迫度 | (Immediacy of the concern) |

•懸念事項を検討するにあたっての本報告書の参考文献

•人権と倫理(Human rights and ethics)

- C. Heyns, 2013, Report of the Special Rapporteur on extrajudicial summary or arbitrary executions, United Nations Human Rights Council, A/HRC/23/47,
- www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/A-HRC23-47_en.pdf.

•合法性(Legality)

- A wide variety of views on legal interpretations area available. See, for example, International Committee of the Red Cross,2016, Autonomous Weapon Systems: Implications of Increasing Autonomy in the Critical Functions of Weapons, pp. 79?82,<https://www.icrc.org/en/publication/4283-autonomous-weapon-systems#>;
- Human Rights Watch and International Human Rights Clinic, 2015, Mind the Gap: The Lack of Accountability for Killer Robots, Human Rights Watch, https://www.hrw.org/sites/default/files/reports/arms0415_ForUpload_0.pdf;
- Kenneth Anderson and Matthew Waxman, 2013, Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can, Hoover Institution, http://media.hoover.org/sites/default/files/documents/AndersonWaxman_LawAndEthics_r2_FINAL.pdf.

4-4 UNIDIR2017年報告書が示す懸念事項の概要

① 人権と倫理

- LAWSに関する国際的な議論の初期の段階では、人権と法的な問題を中心とした議論が中心
- 議論が進むにつれて、政府が自律性を有する兵器システムの開発を実際に進展させているため、当初とは異なる問題と懸念事項の深刻化が顕著になりつつある

倫理的側面における問題

- 1899年第2次ハーグ陸戦条約の前文
- 1977年追加議定書IIIに定められたマルテンス条項
- 「実効性を有する法律が適用されない場合であっても、人は人道の原則に基づいて保護され、公共の良心に従う。」

UNIDIRにおける倫理的問題の検討

- 人間の尊厳に対し、自律性を有するシステムがその生死の判断を自律的に行うことについて、倫理的な側面からの検討が不可欠であると指摘
- 国際人道法では倫理的課題について規制する法律が存在しない場合においても、マルテンス条項第4条の解釈に基づいて対応
- CCWにおける検討は、国際人道法をどのように遵守するのか技術的な側面からも検討が必要であると指摘
- UNIDIR, 2015, *The Weaponization of Increasingly Autonomous Technologies: Considering Ethics and Social Values*, UNIDIR Resources no. 3<
<http://www.unidir.org/files/publications/pdfs/considering-ethics-and-social-values-en-624.pdf>>. LAWSとマルテンス条項の関係についての考察として、Peter Asaro, *Jus nascendi, Robotic Weapons and the Martens Clause*, ROBOT LAW 367 (Ryan Calo et al. eds., 2016).

② 合法性

- 追加議定書1の36条に基づく新しい兵器や戦争手段の検討にあたって、当該システムに組み込まれるアルゴリズムのテストについて、それらの使用にあたっての法的な観点からの説明責任や武力行使における人間の役割と責任について検討が必要であると指摘
- 合法性の観点からの検討においては、技術的問題と法的問題は相互に密接に関係するものであるため、技術開発において合法性の観点からの基準をどのように満たすことができるのか検討が必要であるとしている

③ 技術

- 技術開発は、その開発の速度と取り組みが異なるだけでなく、技術者と政府の考え方や評価も異なる
- 純粋に技術的な評価に関して政策的な議論を同じ位置で行うことは困難

技術的な懸念事項

- 予測可能性や信頼性の確保のために求められる方法
- 意図しないリスクの軽減方法
- 追加議定書36条が定める検証に至るまで多岐にわたる

④ 拡散及び軍拡競争

- 過去3年間のCCWの非公式専門家会議における議論では、自律兵器の拡散の危険性が指摘されてきている
- 小型で効率的なシステムを開発することが可能である現状
- 拡散を完全に制限することは困難
- テロリストなどの悪意ある者やグループによる利用を防ぐことは困難
- 軍事技術の開発においては技術的な優位性を維持することが重要
- 自律性を向上させるインセンティブがある
- 自律兵器そのものの開発のみならず、その制御システムへの不正アクセスなどの技術開発や強力な電磁兵器の開発へのインセンティブも高まる

4-5 UNIDIR2017年報告書が示す懸念事項の概要

⑤ 戦略的安定性

- 自律兵器の拡散によって世界的に不安定な状況が生ずること
- 武力行使の閾値が下がる可能性があること
- 戦場における自律性の向上によって瞬時に作戦が実行され完了する「フラッシュ戦争(Flash War)」の勃興による不安定さの助長への懸念

⑥ 技術のデュアルユースの側面

- デュアルユースの問題は自律兵器をめぐる問題に限られない
- 化学兵器禁止条約、生物毒素兵器禁止条約などの枠組みにおいてデュアルユースに係る技術や物資の規制がなされてきた
- ミサイル技術管理や核不拡散さらにはワッセナーアレンジメントなどにおける取り組みがあげられる
- 自律兵器開発においては、その研究開発を制限することは、民間部門における自律性を有する技術の研究開発の支障になりかねないと指摘する政府もあること

⑦ 運用上の懸念

- 自律兵器についての差し迫った運用上の懸念
- システムが意図した機能を有するか、信頼性や予測可能性については不確実な要素がある実戦において利用するには懸念を払拭できる状況にはないと指摘

⑧ 意図しないリスクと安全性の問題

- 複雑かつ危険な技術には意図しないリスクが伴うこと
- 設計者や利用者が意図しない問題を引き起こす可能性
- 人命に対する脅威はコントロールできない可能性やリスクが存在
- 高度に自動化されたシステムは予測不可能な動作をすることがある点
- 高度なシステム設計は慎重になされ冗長性も確保されてはいるものの、そのような問題が発生することを全て防ぐことは困難であることを指摘

⑨ 「静的」又は「動的」自律性

- CCWIにおける議論は、もっぱら、動的な兵器システムをめぐる問題に焦点を当てている自律性を有する軍事技術については、意思決定支援などの静的なシステムについても見落としてはならないと指摘
- 目標の選択や戦闘における作戦の決定など軍事力を行使する際の決定をサポートする際に用いられるシステム(そのようなシステムは、直接弾丸を発射するシステムではない)

⑩ 懸念の切迫度

- 現段階で自律兵器と考えられるものは、既存のロボット兵士のスマート版にすぎない
- 将来的な技術革新により、当該兵器は長期間にわたって通信や充電をしなくても自律的に動作し続けるようになる
- ターミネーターの脅威といったシナリオで問題を考えることは容易ではあるが、短期的かつ潜在的な開発を懸念するのか、それとも遠い将来の開発も含めて問題を考えるべきなのか喫緊かつ緊急性が高い問題であることを認識すべきであると指摘

4-6 用語の適切な認識の必要性（報告書において確認すべき事項とされているもの）

誤解に基づく議論がなされているため、明確な定義と用語の適切な認識が必要

自動化と自律性

(Automation or autonomy)

学習、適応および調整

(Learning, adaptation and adjustment)

最適化

(Optimizing)

精査性/説明性

(Scrutability/explainability)

致死率

(Lethality)

予測可能性と信頼性

(Predictability and reliability)

精度と正確性

(Precision and accuracy)

説明責任

(Accountability)

人間の決定とそれに依存しない機械

(Human decisions and agnostic machines)

システムのシステム

(Systems of systems)

通信とコネクティビティ

(Communications and connectivity)

モビリティ

(Mobility)

速度

(Speed)

技術中心のアプローチ (Technology-centric approach)

人間中心アプローチ (Human-centred approach)

タスク・機能アプローチ (Task/Functions approach)

アプローチの順序付け (Sequencing the approaches)

- 類似の定義付アプローチをしている文献
- Vincent Boulanin, *Mapping The Development of Autonomy in Weapon Systems: A Primer on Autonomy*, SIPRI, 2016, pp. 29-30
- <https://www.sipri.org/sites/default/files/Mapping-development-autonomy-in-weapon-systems.pdf>

4-8 各国政府や国際機関によるLAWS定義の試み

オランダ政府

- Autonomous Weapon System: “A weapon that, without human intervention, selects and engages targets matching certain predefined criteria, following a human decision to deploy the weapon on the understanding that an attack, once launched, cannot be stopped by human intervention.”
- 参考文献
 - Government of the Netherlands, 2017, “Examination of Various Dimensions of Emerging Technologies in the Area of Lethal Autonomous Weapons Systems, in the Context of the Objectives and Purposes of the Convention”, document CCW/GGE.1/2017/WP.2 of 9 October.

フランス

- “Lethal autonomous weapons are fully autonomous systems. [...] LAWS should be understood as implying a total absence of human supervision, meaning there is absolutely no link (communication or control) with the military chain of command. [...] The delivery platform of a LAWS would be capable of moving, adapting to its land, marine or aerial environments and targeting and firing a lethal effector (bullet, missile, bomb, etc.) without any kind of human intervention or validation. [...] LAWS would most likely possess self-learning capabilities.”

赤十字国際委員会

- Autonomous Weapon System: “Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.

スイス政府

- Autonomous Weapon Systems: “Weapons systems that are capable of carrying out tasks governed by IHL in partial or full replacement of a human in the use of force, notably in the targeting cycle.”

イギリス

- “An autonomous system is capable of understanding higher-level intent and direction. From this understanding and its perception of its environment, such a system is able to take appropriate action to bring about a desired state. It is capable of deciding a course of action, from a number of alternatives, without depending on human oversight and control, although these may still be present. Although the overall activity of an autonomous unmanned aircraft will be predictable, individual actions may not be.”
- Article 36, 2016, “The United Kingdom and Lethal Autonomous Weapons Systems: Analysis of UK Government Policy Statements on Lethal Autonomous Weapons Systems”, <http://www.article36.org/wp-content/uploads/2016/04/UK-andLAWS.pdf>.
- United Kingdom Ministry of Defense, 2017, Unmanned Aircraft Systems, Joint Doctrine Publication 0-30.2, p. 43,
- https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/640299/20170706_JDP_0-30.2_final_CM_web.pdf.

米国

- “A weapon system that, once activated, can select and engage targets without further intervention by a human operator. This includes human-supervised autonomous weapon systems that are designed to allow human operators to override operation of the weapon system, but can select and engage targets without further human input after activation.”

CCWの2018年報告書

- Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, CCW/GGE.1/2018/3, 23 October 2018.

同2019年の報告書

- Group of Governmental Experts of the High Contracting Parties to the Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, CCW/GGE.1/2019/CRP.1/Rev.2, 21 August 2019.

報告書の位置づけ

- 2019年の委員会では、2018年に検討がなされた原則について引き続き検討を行うことについて確認
- 継続的な検討にあたっては、5つの議題に基づいて追加で検討が必要な事項について討議2019年の報告書では、ヒューマン・マシン・インタラクションについての原則が新たに追加
- 国際人道法に対する自律型兵器などの新たな技術によってもたらされる潜在的な課題の探索に関する議題が追加
- CCWでは、原則案が提示されてはいるものの、当該原則案は確定したものではなく継続的に検討を行う原則
- 原則案についてはさらに検討を進めること、法的、技術的及び軍事的側面における検討を継続することを提示
- 2019年の報告書では2017年、18年、19年の報告書を反映した結論を示すにとどまっている

構成

- 新たな共通点、結論および提言を示し、三つの提言から構成される
- 作業計画(CCW / GGE.1 / 2018/2)に基づいて検討に着手し、以下の議題に係る項目について検討を実施。

目的

- (a) 致死性自律兵器システム分野における新興技術によって国際人道法にもたらされる可能性がある潜在的な課題の調査(2019年報告書で追加)
- (b) 条約の目的及び当該目的に関連する概念及び特性に関する共通の理解を促進するため、検討対象のシステムの特徴を明確にすること。
- (c) 致死兵器の使用における人的要素の継続的検討。自律型致死兵器システム領域における新技術の開発、展開及び使用におけるヒューマン・マシン・インタラクションの側面を中心に。
- (d) 当グループの検討課題との関連性において関係する技術が将来的に軍事利用される可能性の検証。
- (e) 自律型致死兵器システム領域における新たな技術によってもたらされる人道及び国際安全保障における検討課題に対処するための提案が、本条約の目的および当該目的の趣旨において、事前審査による政策決定や過去、現在および未来の提案に左右されることなく選択可能な提案の検討が行われること。

5-2 原則指針の提案

原則指針提案の前提

- 国際法、特に国際連合憲章及び国際人道法 (IHL) 並びに関連する倫理的問題に関し、当グループにおける検討を継続して行うことを確認
- 自律型致死兵器領域における新たな技術によって将来的にもたらされる可能性がある課題が、国際人道法にもたらす影響を踏まえ、以下のことを確認するとともに、今後の議論はこれに左右されるものではないことを確認

原則指針の意義

- CCW/GGEの報告書の中核をなす部分
- LAWSの規制に向けた検討の礎となる基本原則の提案

原則指針案

- (a) 国際人道法が、自律型致死兵器システムの将来的な開発と使用を含め、すべての兵器システムに完全に適用され続けること。
- (b) 兵器システムの使用に関する決定における人的責任を保持し、アカウントビリティを機械に委譲してはならない。これは、武器システムのライフサイクル全体を通して考慮されなければならない。
- (c) CCWの枠組みにおいて新たな兵器システムを開発、配備及び使用する際の責任は、人間の指揮統制の責任の範囲内で当該システムを運用することを含め、適用できる国際法に基づかなければならない。
- (d) 新たな兵器、手段又は戦闘方法の研究、開発、取得又は採用において、その使用の一部又は全部が国際法によって禁止されるか否かにかかわらず、国際法の下における国家の義務に従う決定がなされなくてはならない。
- (e) 自律型致死兵器システムの領域において、新技術に基づく新しい武器システムを開発し又は取得する場合、物理的セキュリティ、適切な非物理的保護手段(ハッキングやデータのなりすましに対するサイバーセキュリティを含む)、テロ集団による取得のリスク及び拡散の危険性を考慮しなければならない。
- (f) リスクアセスメント及びその低減策は、あらゆる兵器システムにおける新技術の設計、開発、テスト及び展開サイクルに組み込まなければならない。
- (g) 自律型致死兵器システム領域において新たに開発された技術を使用するにあたっては、国際人道法及びその他の適用される国際法上の義務が遵守されるよう検討しなければならない。
- (h) 将来的な政策を策定する際に、自律型致死兵器システム領域において新たに開発される技術を擬人化してはならない。
- (i) CCWが関係する議論及び将来的な政策判断は、インテリジェントな自律技術の平和的な使用の促進又はアクセスを妨げてはならない。
- (j) CCWIは、本条約の対象と目的の観点から、自律型致死兵器システム領域において新たに開発された技術と関係する問題に対処するための適切な枠組みを提供し、軍事的必要性と人道的配慮のバランスを追求する。

原文

- It was affirmed that international law, in particular the United Nations Charter and international humanitarian law (IHL) as well as relevant ethical perspectives, should guide the continued work of the Group. Noting the potential challenges posed by emerging technologies in the area of lethal autonomous weapons systems to IHL, the following were affirmed, without prejudice to the result of future discussions:
- Possible Guiding Principles
 - (a) International humanitarian law continues to apply fully to all weapons systems, including the potential development and use of lethal autonomous weapons systems.
 - (b) Human responsibility for decisions on the use of weapons systems must be retained since accountability cannot be transferred to machines. This should be considered across the entire life cycle of the weapons system.
 - (c) Accountability for developing, deploying and using any emerging weapons system in the framework of the CCW must be ensured in accordance with applicable international law, including through the operation of such systems within a responsible chain of human command and control.
 - (d) In accordance with States' obligations under international law, in the study, development, acquisition, or adoption of a new weapon, means or method of warfare, determination must be made whether its employment would, in some or all circumstances, be prohibited by international law.
 - (e) When developing or acquiring new weapons systems based on emerging technologies in the area of lethal autonomous weapons systems, physical security, appropriate non-physical safeguards (including cyber-security against hacking or data spoofing), the risk of acquisition by terrorist groups and the risk of proliferation should be considered.
 - (f) Risk assessments and mitigation measures should be part of the design, development, testing and deployment cycle of emerging technologies in any weapons systems.
 - (g) Consideration should be given to the use of emerging technologies in the area of lethal autonomous weapons systems in upholding compliance with IHL and other applicable international legal obligations.
 - (h) In crafting potential policy measures, emerging technologies in the area of lethal autonomous weapons systems should not be anthropomorphized.
 - (i) Discussions and any potential policy measures taken within the context of the CCW should not hamper progress in or access to peaceful uses of intelligent autonomous technologies.
 - (j) The CCW offers an appropriate framework for dealing with the issue of emerging technologies in the area of lethal autonomous weapons systems within the context of the objectives and purposes of the Convention, which seeks to strike a balance between military necessity and humanitarian considerations.

5-3 条約の目的と検討対象のシステムの特徴

条約の目的及び当該目的に関連する概念及び特性に関する共通の理解を促進するため、検討対象のシステムの特徴を明確にすること。

1. 一部の国家にとっては、自律型致死兵器システムの実用的な定義は、問題が指摘されている潜在的なリスクに適切に対処するために不可欠である。その他の国家にとっては、定義に関する合意がないことによって、CCW内の議論やその発展を妨げるべきではない。特徴の明確化又は実用的な定義を定めるにあたっては、その方針を事前に決定することや偏見に基づくことなく政策を決定しなければならない。また、利害関係者によって普遍的に理解されるものでなければならない。
2. 技術が急速に進化していることに鑑み、物理的性能、耐久性又は目標捕捉及び交戦段階における精巧さといった純粋に技術的な特徴だけでは、自律型致死兵器システムの特徴を明確にするには不十分である可能性が高い。
3. 機械学習(外部から提供された学習データなし)及び自己進化(人間による設計入力なし)に関連する技術的特徴は、さらなる研究がなされる必要がある。同様に、技術的基準のみに基づいて自律性の一般的なしきい値レベルを定義しようと試みることは、自律性がスペクトルであり、その理解が技術フロンティアの変化と共に変化し、武器システムの異なる機能が異なるレベルの自律性を有する可能性があるため、定義を明確にすることが困難になる可能性が高い。
4. GGEのマandatにおいて明示された致死性は、敵対行為(the conduct of hostilities)に関連するすべての規則の適用及び尊重を害するものではない。
5. 軍事目標の捕捉及び交戦サイクルにおける自律性は、自律性が目標捕捉の全部又は一部に存在する可能性があり、近接戦闘などの他の状況ではさらに適用される可能性があることを考慮して、さらなる研究を行う必要がある。
6. CCWの趣旨に基づき、兵器の使用及び機械とのインタフェースとの関係における人的要素に関連する特性に焦点を合わせることは、説明責任(アカウンタビリティ)と責任の問題に対処する上で必要である。

5-4 致死性兵器の使用における人的要素

- 自律型致死兵器システム領域における新技術の開発、展開及び使用におけるヒューマンマシン・インタラクションの問題
- CCWの目的と趣旨に基づき、自律型致死兵器システム領域における新技術の開発、展開及び使用におけるヒューマン・マシン・インタラクションの性質や特性をめぐる問題について検討を行うことが重要であること

0) 開発前段階における政治的指向 (Political direction in the pre-development phase)

1) 研究開発 (Research and development)

2) 試験、評価及び認証 (Testing, evaluation and certification)

3) 展開、訓練、指揮統制 (Deployment, training, command and control)

4) 使用及び使用の中止 (Use and abort)

5) 使用後の評価 (Post-use assessment)

- 人的要素(コントロール)については、赤十字国際委員会の報告書(ICRC, Autonomy, artificial intelligence and robotics: Technical aspects of human control, 20 August 2019 <<https://www.icrc.org/en/document/autonomy-artificial-intelligence-and-robotics-technical-aspects-human-control>>)においては、民間部門における自律型システムの利用における人間によるコントロールをめぐる経験を生かすべきであると指摘している
- ヒューマン・コントロールに焦点を当てた考察として、Thompson Chengeta, *Defining the Emerging Notion of "Meaningful Human Control" in Weapon Systems*, 49 N.Y.U. J. Int'l L. & Pol. 833(2017).

ヒューマンマシン・インタラクションに係る検討事項

- 1. 説明責任は、CCWの趣旨に基づき様々なヒューマン・マシン・タッチポイントにおいて一貫する問題である。人間は、武力行使に関する決定において適用される国際法に従い、常に説明責任を負う。
- 2. 独立した倫理審査を通じて、国家安全保障上の考慮事項や営業秘密の利用に関する制限を念頭に置いた学際的な検討を研究開発において実施することが望ましい。
- 3. 開発中の兵器システム又は既存の武器システムの使用を大幅に変更する改造は、国際人道法を確実に遵守した上で実施しなければならない。
- 4. あらゆる可能性又は想定される意図を網羅した使用シナリオの検証可能性及び認証手順を開発するにあたっては、そのような手順の適用の経験に関し、国家安全保障上の考慮事項又は営業秘密に関する規制を考慮して共有しなければならない。
- 5. 武力紛争における武力行使の説明責任は、責任ある指揮統制の範囲内における新たな兵器システムの運用を含め、適用される国際法に従わなければならない。
- 6. 武力行使における人的責任は保持されなければならない。この点について、国際人道法の遵守を確実にするために、武器使用の介入に拡張することが可能な限り求められる。
- 7. 国際人道法を遵守し、新たな技術の開発と展開のサイクルを通して人間による説明責任と責任を保持するために、人的資源及び訓練に必要な投資がなされるべきである。
- 8. 以上を踏まえ、この分野における国家の権威と責任を認識しつつ、兵器システムのライフサイクルの各段階において、これらのフェーズ全体を通して説明責任を明確にすることと同様に、人間と機械の相互作用の程度と質について共通の理解を得ることについて議論を続けることが有用である。

6. LAWS反対運動の展開

「キラー・ロボット・ストップキャンペーン*」

- Campaign to Stop Killer Robots
 - <https://www.stopkillerrobots.org/>
- 2007年8月にNoel Sharkeyが完全自律型致死ロボットの開発への懸念を表明する意見*を公開したことに端を発し、2012年10月からNGOとして活動
 - Noel Sharkey, Robot wars are a reality: Armies want to give the power of life and death to machines without reason or conscience, Sat 18 Aug 2007
<https://www.theguardian.com/commentisfree/2007/aug/18/comment.military>

NGO「ロボット軍備管理国際委員会 (the International Committee for Robot Arms Control(ICRAC))」

- ICRC
 - <https://www.icrac.net/>
- Noel Sharkey, Jurgen Altmann, Peter M. Asaro, Robert Sparrowによって2009年からNGOとして活動。
- 武装自律型無人システムの開発、配備及び使用の禁止をMission Statementとして提言した
 - ICRAC, Mission Statement, September, 2009 .
<https://www.icrac.net/statements/>

「失われつつある人間性: 殺人ロボットに反対する根拠」

- Human Rights Watch, Losing Humanity: The Case against Killer Robots
 - <https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots>
- 「殺人ロボット: 完成前に禁止を 完全に自律稼動する兵器 増える市民への危険」2012年11月19日
 - <https://www.hrw.org/ja/news/2012/11/19/248112>

今、考えていること。これから考えたいこと

技術の進化は誰にも止めることはできない(技術革新を阻んではいけない)

反対運動はCCWや各国政府における取り組みに一定の影響力を及ぼすことは可能であっても、新興技術の利用をテロリストや反社会的な勢力に思いとどまらせることはできない実効性及び強制力なき単なる反対論や、空虚な規制論に陥ることは避けたい

CCWの2019年版の報告書では、2018年に示された基本原則の提案を発展させる必要性を議論することで、今後の技術動向の変化に柔軟に対応できる原則となっている

非公式専門家会議から政府専門家会合の議論においても、国際法、特に国際連合憲章及び国際人道法(IHL)並びに関連する倫理的問題に関し、CCWの枠組みにおける検討を継続して行うことを「確認」するにとどまっている

自律型の致死兵器システム関連領域における新たな技術によって将来的にもたらされる可能性がある課題が、国際人道法にもたらす影響を踏まえ必要な確認事項についても検討は行いつつも、将来的にどのように自律型システムに係る技術に依拠するLAWSが進化するのか見極めることが困難であることは否定できない

今後の議論の展開は、現在までの検討の範囲に制限されるものではないことは各国政府共通の認識とする段階にまでは至っていることは、LAWSの規制のあり方に関する議論をさらに深化させることが可能であることを意味する

今、考えていること。これから考えたいこと

新興技術の開発から利用に至る展開は、軍事技術として開発された技術が民間部門において利用される場合と、民生技術が軍事目的で用いられる場合がある

新たな技術開発には多額の研究資金が必要となることから、予算額が大きい軍事予算を背景とした研究が、結果的に民間における技術革新を促進させてきた事実があることも否定できない

例えば、インターネットの発展は、国防総省の国防高等研究計画庁(DARPA)が研究予算を投じて開発されたARPAネットを原型とすることはよく知られている。カーナビや様々な位置情報を利用するサービスにおいて用いられているGPSも同様

自律型兵器の分野においては、兵器開発における自動化や遠隔操作の技術はもとより、AIやドローンなど民間において研究開発が進んできたものを、新たな自律性を有する軍事技術として発展させようというもの

軍需を背景とする軍事技術の開発によって将来的に民間において利用することができる技術がどの程度あるのか、現時点では定かではないが、その点が、いわゆるロボット兵器の規制を主張する論拠として、ロボット兵器の研究開発を規制したとしても将来的に民間における技術開発が制限されることがないという考えの根底にあるのかもしれない

技術革新が進めば進むほど、リスクも必然的に増大する

- そのリスクについて、①利用することの脅威、②利用しないことによる脅威、③利用に反対することによる脅威の三つの側面から考えるべきではないだろうか

①利用することの脅威

- AIそのものへの脅威、AIを用いたサービスや装置の利用に伴う脅威、AIを利用した犯罪やテロなどである。

②利用しないことによる脅威

- 軍事バランスのように安全保障上のバランスが崩れることによる脅威、テロリストなど正規軍による利用ではなく法的にも国際法上も制限の対象外における利用がなされる脅威などがあげられる。

③利用に反対することによる脅威

- 理想論としても、AI兵器の開発は人類への脅威になることが明らかであるが故に、絶対的に禁止すべきであると考えられる
- しかし、技術の進化は誰にも止めることができず、その利用を止めることも困難である
- 核不拡散防止のように、核開発といった高度な技術を要するものとは異なり、市販の数万円のドローンと武器の組み合わせによって、簡単に強力な兵器を作り出すことができる状況で、それらにAIが搭載された自律型兵器の開発を禁止したり停止されることは事実上不可能であろう
- 利用に反対することは簡単であるが、実際に研究開発や利用を禁止することは極めて困難な状況になりつつあるため、安全保障をめぐる問題においてAIや自律型兵器をめぐる問題を議論しないことは、将来的に禍根を残す最大のリスクである