人工知能倫理・ガバナンス及び持続可能な開発
翻訳シリーズ
Translation Series on Artificial Intelligence Ethics, Governance, and Sustainable Development

人工知能倫理とガバナンスにおける異文化協力の障壁の克服
Overcoming Barriers to Cross-cultural Cooperation in AI Ethics and Governance

北京知源人工知能研究所人工知能倫理とセキュリティ研究センター
中国科学院自動化研究所中英人工知能倫理とガバナンス研究センター

Research Center for AI Ethics and Safety, Beijing Academy of Artificial Intelligence
China-UK Research Centre on AI Ethics and Governance, Institute of Automation, Chinese Academy of Sciences

前書き

人工知能の倫理とガバナンスは、全世界の人工知能の発展と革新の方向と将来に関係している。 人工知能を人間・社会・生態及びグローバルな持続可能の発展を促進するための技術として使用することは、人工知能技術の革新に対する人類共通のビジョンであろう。このプロセスにおいては、各国及び政府間組織および国際組織による人工知能の倫理とガバナンスは、学術機関・産業・政府を通じてさまざまな方法で関連する原則・方針・基準・法律の制定・技術の着陸を積極的に推し進めている。各国・組織は異なる文化背景の下で努力してきたものであるが、「論語」が言う「君子は和して同ぜず」のように、文化上の差異はものを考える時の異なる視点と、互いから学ぶ機会を、私たちに提供できる。異文化間における相互信頼の確立は、世界の調和のとれた発展の礎石である。人工知能の倫理・ガバナンス・持続可能な開発は、世界の科学・技術・社会の分野における持続的かつ重要な議題になる。このため、北京知源人工知能研究所の人工知能の倫理及びセキュリティ研究センターは、中国科学院自動化研究所の中英人工知能の倫理及びガバナンス研究センター等の機関と協力し、「人工知能倫理・ガバナンス・持続可能な開発　翻訳シリーズ」を立ち上げ、人工知能の倫理とガバナンス、持続可能な開発の分野における重要な文献を選考し、翻訳を整理し、世界中の読者に紹介する。異なる文化、言語の交流から互いに貢献し合い、技術開発に伴う文化交流を促進し、世界の人工知能と人類の将来との調和の取れた発展を推進することが期待されている。

曽毅
北京知源人工知能研究所人工知能倫理とセキュリティ研究センター　センター長
中国科学院自動化研究所中英人工知能倫理とガバナンス研究センター　センター長

The ethics and governance of Artificial Intelligence (AI) are essential for the direction and future on the development and innovation of global Artificial Intelligence. Using AI as an enabling technology to promote the sustainable development of humanity, society, ecology are the common vision for the global technology innovation of AI. In this process, the ethics and governance of AI from various countries, intergovernmental organizations and international organizations actively promote the development of relevant principles, policies, standards, laws, and their technology and society groundings through academic institutions, industries, governments, etc. Although the efforts of various countries and organizations are established in different cultural contexts, cultural differences provide us with different perspectives and opportunities to learn from each other. As the analects by Confucius say, "Be in harmony, yet be different". Building cross-cultural mutual trust is the foundation of global harmonious development. AI ethics, governance and sustainable development will continuously be an important topic in the advancement of science, technology and society. Hence, the research center for AI Ethics and Safety, Beijing Academy of AI, together with the China-UK Research Centre for AI Ethics and Governance at the Institute of Automation, Chinese Academy of Sciences, jointly launched the "Translation Series on AI Ethics, Governance and Sustainable Development". Efforts will be put to selection and translation of important documents in the field of AI Ethics, Governance and Sustainable Development, and introduction of them to readers around the world. We look forward to cross-cultural and cross-language exchanges, and promoting the harmonious development of AI for the world, for humanity and for the future.

Yi Zeng
Director, Research Center for AI Ethics and Safety, Beijing Academy of Artificial Intelligence
Director, China-UK Research Centre for AI Ethics and Governance, Institute of Automation,
Chinese Academy of Sciences

人工知能の倫理とガバナンスにおける異文化協力の障壁の克服[1]

Seán S. ÓhÉigeartaigh[1,2], Jess Whittlestone[1], Yang Liu[1,2], 曽毅[2,3], 劉哲[2,4]
1. ケンブリッジ大学レヴァーヒューム知能未来研究センター，イギリス
2. 中英人工知能倫理とガバナンス研究センター，中国
3. 北京知源人工知能研究所人工知能倫理とセキュリティ研究センター，中国
4. 北京大学哲学と人間未来研究センター，中国

要約

人工知能(AI)のグローバルなメリットの実現には、多様な文化的視点と優先事項を考慮しながら、ガバナンスと倫理基準の多くの分野での国際協力が必要である。 現在、これを実現するには、異文化間におけ信頼の欠如や、地域を超えた協力というより現実的な課題など、多くの障壁がある。ヨーロッパと北米、そして東アジアとの間の協力は、人工知能の倫理とガバナンスの発展に大きな影響を及ぼしているため、この論文では主に協力の過程で上記した地域が直面している障壁に焦点を当てている。

この論文では AI の倫理とガバナンスに関する異文化間の協力が拡大していくことに対して楽観的になる理由があると信じている。ただ基本的な意見の相違と比較すると、文化や地域間の誤解は、異文化間の信頼を損なう上で一般に考えられているよりも、より深刻な影響があると考えている。もっとも基本的な違いが存在するとしても、必ずしも次の 2 つの理由で異文化間の実りある協力を妨げるものではない。その原因は以下のごとくである。(1) 協力には AI のすべての領域において原則と標準に関する合意を達成する必要はない。(2) より抽象的な価値観や原則についての意見の不一致があるとしても、実際的な問題について合意に達する場合が時々ある。AI の倫理とガバナンスにおける異文化協力の促進において、相互理解のための強固な基盤を築き、必要かつ可能である場合は差異を確認しながら共通の立場を求めることを明確にすること、この点において学界は重要な役割を果たしていると考えている。そこでこの論文では、主要な文書の翻訳と多言語公開、研究者交流プログラム、異文化トピックに関する研究課題の推進など実践的なアクションと計画について、いくつかの推奨事項を提案する。

キーワード：人工知能；人工知能倫理；人工知能ガバナンス；異文化協力

[1] この論文の英語版は、2020 年 5 月 15 日に Philosophy and Technology Journal でオンラインで公開された。元の英語版は、次の URL でアクセス可能である：https://link.springer.com/article/10.1007/s13347-020-00402-x。この論文の中国語訳は中国科学院自動化研究所中英人工知能倫理とガバナンス研究センター(http://www.ai-ethics-and-governance.institute/)によって翻訳された。担当著者である Seán S. ÓhÉigeartaigh の連絡先は（so348@cam.ac.uk） であり、2 人の国内の著者である曽毅と劉哲はそれぞれ(yi.zeng@ia.ac.cn) と(liu.zhe@pku.edu.cn)である。

1. はじめに

人工知能は広範にわたって使用されているため、世界中の多くの国にとってコアテクノロジーとみなされている (Brynjolfsson and McAfee, 2014)。AI 技術、特に機械学習技術は、言語翻訳・科学研究・教育・ロジスティクス・輸送など、幅広い多くの分野に効果的に適用されている。国内、国際、またはグローバルな観点から、人工知能は明らかに経済・社会・文化に大きな影響を与えている。 その結果、AI 倫理や AI ガバナンスに対する注目は日に日に高まってきている。AI 倫理とは、AI システムが人間の幸福や他の定着した価値(自律性や尊厳など)に潜在的な影響を与えていることに鑑み、それをどのように開発及び展開すべきかについての質問を指す。AI ガバナンスの問題は実践性とより密接に統合されていて、基本ルール・ガバナンスフレームワーク・またはより「柔軟な」方法(業界規範や倫理規定など)を通じて、社会における AI の適用が倫理的であることを確保することを指す[2]。

異文化協力は、関連する倫理およびガバナンスの取り組みを成功させるために不可欠である。ここでの「異文化協力」とは、具体的には AI 技術の開発・応用・ガバナンスが社会に利益をもたらすことを確実にするための、異なる文化的背景または国のグループの協力を指す。 この論文では主に国境を越えた協力について焦点を当てている。具体的な例には以下が含まれる(ただし、これらに限定されない):さまざまな国の AI 研究者が協力してプロジェクトを完了し、安全かつ信頼できる方法で AI システムを開発する。AI の倫理的問題に焦点を当てた国際的な議論は同様に多様な国際的な視点を利用できるように、さまざまなコミュニケーションチャネルを確立する。実践的なガイドライン・基準・規制の策定に参加するようにすべての国の利害関係者を招待する。異文化協力を奨励する必要があるが、これは必ずしも世界のすべてが AI に関する同じ規範・基準または規制に従うものではなく、そしてすべての方面において国際協定を締結する必要があることを意味するわけではない。 グローバルな基準や合意が必要な問題、さらに文化的な違いが必要な場所を特定することは、それ自体が重要な課題であり、それらを解決するための協力が必要である。

異文化協力の重要性は、次の点に反映されている。まず、AI はグローバルな社会的利益をもたらし、ある地域の高度な技術を共有することにより他の国の発展を促進し、社会が共に進歩し、そして様々な地域に一貫した前向きなメリットを確保するため、上記の目標を達成するために協力は不可欠である。 次に、その協力により世界中の研究者が専門知識・リソースおよび実例を共有できる。それにより、有益な人工知能の適用を早期に実現でき、潜在的な倫理上の問題や主要な安全上の問題も合理的に処理できる。 第三に、適切な協力の欠如は、国または異なるビジネスエコシステム間の競争圧力が安全で、倫理的で、および社会的に有益な AI 開発への投資不足につながるリスクがある (Askell et al. 2019; Ying 2019)。最後に、国際協力の重要性は多くの実際的な要因にも反映されている。たとえば、国や地域の境界を越える人工知能アプリケーション(主要な検索エンジンやインテリジェント運転での応用など)をさまざまな規制環境に効果的に統合させ、 そして他の地域との技術的な相互接続を実現させることを確保する場合に反映されている (Cihon 2019)。

[2] ガバナンスプログラムは通常倫理原則の実際の具体化であり、倫理フレームワークは関連するポリシーと規制の開発の出発点でもあるため、人工知能倫理と人工知能ガバナンスは密接に関連している。今は、人工知能倫理はより実践的であり、社会における人工知能の適用に関する多くの倫理原則およびガイドラインが派生されているため、多くの「ソフト」ガバナンスプログラムも人工知能倫理に組み込まれている。従って、この章の「人工知能の倫理とガバナンス」とは一般に、倫理的問題、帰納的問題を判断し、そしてガバナンス手法に適用するプロセス全体を指す。

東アジアとヨーロッパからの主要な学者グループの洞察[3]に基づいて、私たちは異文化協力がAI 倫理とガバナンスにおいて直面する障壁について詳細に分析し、実現可能な解決策を提案する。下記の地域は現在、AI の倫理とガバナンスに関する国際的な対話において極めて重要な役割を果たしているため、この論文はヨーロッパ、北米、東アジアの間の協力に焦点を当てている(Jobin et al 2019)。最近、AI の開発とガバナンスの領域におけるこれらの地域の国家間、特に中国と米国の間の競争と矛盾について、多くの分析が書かれている。この論文での議論と提案は、人工知能をめぐる幅広い国際協力に適用され、そしてこれにより、地域間のより多くの協力が促進されることが期待されている。

AI システムは完全に近づくと、その応用はより効果的で一般的になるにつれ、リスクが徐々に高まる。異文化間の誤解や信頼の欠如が学術的および公開の議論に徐々に定着するようになると、長期的な協力関係を確立することはより困難になる可能性がある。これを念頭に置いて、グローバルな文化協力はできるだけ早く確立されるべきと考える。したがって、AI の影響を導くことに関する共通の理解と深い協力関係を育むことは、グローバル社会にとって緊急かつ差し迫った課題と見なされるべきである。

2. グローバル AI 会話の形成における北米、ヨーロッパ、東アジアの役割

特に北米、ヨーロッパ、東アジアは、企業および政府の投資に支えられて、基礎的および応用的な AI 研究開発の両方に多額の投資を行っている (Benaich and Hogarth 2019; Haynes and Gbedemah 2019; Perrault et al. 2019)。多くの論文では、競争の観点から米国と中国の人工知能の開発と応用の進展を取り上げてきたが (Simonite，2017；Allen, Husain，2017；Stewart，2017)、このようなフレームワークは、規範的で記述的な根拠の両方で批判を受けています (Cave, Ó hÉigeartaigh，2018)。

慎重に検討した結果、これらの地域の学者や政策コミュニティーは積極的に対応し、地域的および世界的レベルの両方から、AI の倫理原則とガバナンスの推奨事項の策定を計画しており、そして政府の関連する取り組みに反映されています。たとえば、人工知能に関する EU ハイレベル専門家グループの活動は、最初の 2 つの出版物の中で倫理ガイドラインとポリシーを公開し、投資の推奨事項を行った[4]。イギリス政府は、「国際的なパートナーと緊密に協力して、人工知能の安全で倫理的かつ革新的な展開を実現する方法について、共通の理解を構築する」（2018 年 5 月）。中国政府はこのような行動をとり、「AI のグローバルなガバナンスに積極的に参加し、ロボットの疎外や安全監視などの主要な国際共通問題の研究を強化し、AI 法と規制、国際ルールに関する国際協力を深める」ことに取り組んでいる（中華人民共和国国務院，2017）。 北米、

[3] 2019 年 7 月に、私たちは異文化間の信頼度に関するシンポジウム(https://www.eastwest.ai/)を開催した。イギリスのケンブリッジ大学とパース大学、中国の香港大学、北京大学、復旦大学および中国科学院、日本の慶応義塾大学と与博古睿研究院中国センターからの代表が参加した。これらの代表は、人工知能の倫理とガバナンスに関する対話、およびヨーロッパ・北米・アジア全体の協力プロジェクトに深く関わってきた。このシンポジウムでは、主に人工知能の異文化間における信頼の構築について、学界の役割が説明された。本論文もそれを参照した。これは、会議に関係する領域のみが重要であるわけではなく、参加者の意見や専門知識は彼らの居住地域のさまざまな意見や専門知識を十分に代表できるという意味でもない。今回のシンポジウムは開発プロセスを促進し、重要なコミュニケーションネットワークを確立し、実現可能性の見解と理論的基盤を提示し、その後の開発のための良い基盤を築いたと信じている。シンポジウムがチャタムハウスルールの下で開催された。

[4] URL：https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

ヨーロッパ、東アジアもそれぞれ、国際標準化機構（ISO）[5]、電気電子技術者協会（IEEE）[6]、経済協力開発機構（OECD）[7]などのさまざまな組織フォーラムで国際 AI 標準化作業の議論を促進するために最善を尽くしている。

北米、ヨーロッパおよび東アジアの顕著な役割は、AI のパートナーシップ (Partnership on AI)[8]、未来の社会 (Future Society)[9]、国際 AI のガバナンス会議 (International Congress for the Governance of AI)[10]および AI のグローバルパートナーシップ (Global Partnership on AI)[11]など、複数の利害関係者や非政府組織に対するリーダーシップと組織に反映されている (Hudson, 2019)。上記の地域では、多くの重要な AI 倫理およびガバナンス会議が開催されている。例えば、米国の「有益な AI」会議シリーズ (US-based Beneficial AI conference series)[12]、北京知源人工知能研究院シリーズ年次総会 (Annual Conferences of Beijing Academy of AI)[13]、北京フォーラム (Beijing Forum)[14]、米国 AI (US-based Artificial Intelligence)、倫理及び社会の会議 (ACM Conference on Artificial Intelligence, Ethics, and Society conference)[15]、先端の機械学習会議から派生したガバナンスと倫理に関するシンポジウムなどである。

この論文では、以下の状況について検討する：

a. 科学技術分野における北米、ヨーロッパ、東アジアのリーダーシップ；
b. グローバルな倫理とガバナンスの対話を推進する上での北米、ヨーロッパ、東アジアの顕著な貢献；
c. 競争の観点から、開発の進展による潜在的な圧力を分析し、倫理とガバナンスの基本的な問題に対する論争の理解統合。

従ってこの論文では、特にこれらの地域や文化間において実りある知的交流と協力を阻む障壁に焦点を当てている。AI 技術の影響範囲が実際にグローバルであることを考慮して、すべての国や文化が果たす役割を検討する必要があるため、この論文では、AI の倫理とガバナンスにおける異文化協力の包括的な分析は行わない。フォローアップの調査では、技術をリードする国とこれらの技術が輸入する国との間であらわれる権力と影響の不平等に注意を払い(Lee，2017)、そして技術輸入国がグローバルなガバナンスと倫理の対話に参加させ、彼らの自律性を高めるための技術の強国の責任に焦点を当てている。

3. 人工知能における異文化協力の障壁

[5] 人工知能標準委員会の 28 か国のうち、22 か国が北米、ヨーロッパ、および東アジア地区からである。(https://www.iso.org/committee/6794475.html)。
[6] 例：電気電子技術者協会 (IEEE)の人工知能原則に関する文献「人工知能デザインの倫理ガイドライン」のメンバーを参考：https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ec_bios.pdf
[7] 例：経済協力開発機構(OECD)と人工知能専門グループ(AIGO)のメンバーを参照：https://www.oecd.org/going-digital/ai/oecd-aigo-membership-list.pdf
[8] https://www.partnershiponai.org/partners/
[9] https://thefuturesociety.org/our-team/
[10] https://icgai.org/icgai-members/
[11] 元国際人工知能事務委員会(International Panel on AI)。カナダとフランスが共同で設立した機構（非人工知能協力機構）(https://www.canada.ca/en/innovation-science-economic-development/news/2019/05/declaration-of-the-international-panel-on-artificial-intelligence.html)
[12] https://futureoflife.org/beneficial-agi-2019/
[13] https://mp.weixin.qq.com/s/tAGOoqqA6ods9uaigWE7uA
[14] http://newsen.pku.edu.cn/news_events/news/global/9133.htm
[15] https://www.aies-conference.com/2020/

AIの倫理とガバナンスに焦点を当てた多くの国際的な同盟関係があったが、AIの開発と応用を導く規範、原則とガバナンスのフレームワークに基づく異文化間協力を本当に実現するには、依然として多くの障壁がある。

異なる地域や文化間の信頼の欠如は、AIの倫理とガバナンスにおける国際協力最大の障壁の1つとなる。現在、米国と中国の学者、技術者、および政策立案者の間では、特に信頼感が欠けている[16]。その理由は次の通り：

(1) 近年、両大国間の政治的緊張の蓄積が深刻になり、AI開発の競争は「東洋」と「西洋」の国家間の競争になっている[17]。

(2) 2つの地域によって提唱された異なる哲学の伝統により、データのプライバシーなどの重要な問題に関して、関連する研究は「西洋」と「東洋」の文化的価値が大きく異なり、相容れないものと考えられている (Larson，2018；Horowitz et al.，2018；Houser，2018)。

最近の技術的および政治的発展もこの不信感を悪化させている。 これには、米国のテクノロジー大手の公的および政治的影響に関する懸念 (Ochigame 2019)、中国の社会信用システムに対する見方と反応 (Chorzempa et al. 2018; Song 2019)、そして論争の的となる分野においてAIテクノロジーの応用に関する懸念が含まれている。その中で広く議論され、論争の的となっている施策には、移民管理のための人工知能の使用 (Whittaker et al.，2018)、米国での犯罪リスク評価 (Campalo et al.，2017)、中国でのウイグル人イスラム教徒グループの追跡 (Mozur，2019)などがある。米国の政治および防衛指導者による敵対的なレトリックは、このような緊張を増大させる。最近のメディアの報道では、(米国が)人工知能領域においての「脅威」となる意図に言及し[18]、より広いレベルで、中国の技術進展が米国の世界的なリーダーシップを脅かすという懸念のため、中国を敵対者として「敵方」と見なされるコメントがあった(2018年6月)[19]。 異文化間の不信感がさらに高まると、人工知能の開発とガバナンスのグローバルな協力の機会が大幅に弱まる恐れがある。

さらに、既存の異文化間の協力と提携がどの程度まで進展したか、そしてそれらが米国、中国、およびこれらの国における大手多国籍企業などの主体性を効果的に規制できるかどうかはまだ不明である。AI倫理フレームワークが複数の利害関係者グループによって原則的に合意できたとしても、それらを実現するために、AIの開発とガバナンスにおいて主導的な役割を果たす当事者の行動を直接制限することは事実上困難である。

グローバルな協力を急ぐ必要がある場合、文化的・地理的要因により、その実施方法に細かい配慮が必要である。しかし、両者のバランスをどうとらえるかという点こそ効果的な協力を実現するためのもう一つの課題である(Hagerty, Rubinov，2019)。1つまたは幾つかの国は単に他の国や地域にその価値観を押し付けようとする状況を回避するよう努めている(Acharya，2019)。特定の領域（医療サービスをサポートするための人工知能の使用など）では、文化によって、利益の

---

[16] この不信感は、7月のワークショップを通じて強調された。 数人の参加者は、中国の参加者が除外された中国のAI進展における地政学的影響に焦点を当てたワークショップに賛同もしくは出席した。「中国に対する取るべき行動」（西洋の観点から）が主要な懸念として提起された。

[17] Ess(2005)は、「東洋」および「西洋」という用語に疑問を投げかけ、国や文化の多様性を表明する正確な用語ではなく、植民地主義の産物であることを示唆した。しかし、より幅広い文化の違いを説明する適切な記述がないため、この2つの用語は様々な文献で広く使用されている。そのため、この論文では2つの用語には制限があると認めつつも、引き続き使用する。

[18] 「多くの人々はAIの脅威について心配しており、私たちは脅威になりたいと思っています。」と米国の国防長官Patrick Shanahanは国防総省の従業員へのメールで述べている (Houser，2018)。

[19] 2019年にワシントンで開催されたセキュリティフォーラムでは、米国国務省の政策立案責任者であるKiron Skinner氏が中国について、「これは、まったく異なる文明とイデオロギーを持つ相手との戦いであり、これまでに経験したことはなく」、「コーカサス人でないコミュニティーでは、そのような強い競争相手に遭遇したのは初めてです」と述べている(Gehrke，2019)。

バランスに対する認識も異なっている(Feldman et al.，1999)。地域固有の標準とガバナンスを実装することは比較的に困難であるが、必要である。AI システムは異なる文化的地域に展開し、異なる効果を生み出すため、異なるガバナンスアプローチを一致させる必要がある(Hagerty、Rubinov，2019)。

AI 開発とガバナンスにおける一部の側面では、協力がより重要になる。例えば、軍事分野での AI テクノロジーの適用（自動化された敵の検出や攻撃など）は、人権や国際人道法の要点に抵触する恐れがある (Asaro，2012)。さらに、自動化された情報収集、意思決定、およびフィードバックの方法が戦場で実装されていると、期待に反し、衝突をエスカレートさせる恐れがある。なぜなら、事件の発生頻度が高くなると、より迅速に対応しなければならないが、人間が適切かつ効果的な判断や監督の実施はより困難になるからである (Altmann，2019)。これらの 2 つの場合では、軍事分野での技術進歩を求める国は AI から恩恵を受けるが、国際的合意や基準がない場合、全体的な影響が極めて不安定になる恐れがある[20]。AI 技術が 1 つの地域で開発され、別の地域で使用または展開される様々の場合では、国際協定が重要である。したがって、異文化協力の主な課題は、国際協定が特に重要である領域を特定し、これらの領域を特定の制約を必要としない他の領域と区別することにある。

国家間協力は、多くの現実的な問題にも直面している。言語の壁、地理的距離、移民制限などにより、異文化間や研究グループ間でコミュニケーションと協力が制限されている。さらに、科学に国境はないとはいうものの、殆どの科学雑誌は英語のままである。

4. 協力における障壁の克服

AI の倫理とガバナンスにおける異文化協力の実現には多くの困難があるが、進展を推し進めるための実行可能な解決策はなお存在し、現時点ではより深い問題を解決する必要はないと考えている。例えば、すべての基本的な倫理的問題と哲学的問題について異文化間のコンセンサスを求める必要はない。そして国家間の数十年にわたる政治的対立を解決する必要もない。

相違点に対する共通認識を含め、より深い相互理解に努める

AI 倫理とガバナンスの将来にとって、国家間の相互不信は深刻な問題である。本論文で言及されている不信感は、一定程度は相互の誤解と誤認による。 そのため、より強固な異文化間の信頼を築くためには、まず正しい認識を構築し、誤解を排除し、異なる文化や国家間の相互理解を強化しなければならない。

AI 分野における東西間の信頼関係を築く際、主な障壁がこれらの地域の価値観の根本的な違いにあることは明白である。したがって、双方は AI に関する開発・応用および AI 倫理のガバナンスについて異なる理解を持ち、時には互いに対立することさえある。異文化間において価値観の相違は確かに存在するが、これらの違いがどのように反映されるかについての主張は、未だ検討されておらず、定見とされた仮説に依存し、実証的な証拠に欠けている (Whittlestone et al. 2019)。「東洋」と「西洋」の倫理的慣習は本質的に矛盾しているという考えは、双方の関係を単純化しすぎている。そこで特定の状況に応じて両地域の内部における異なる哲学的伝統を検討する必要がある。たとえば、中国・日本・韓国においても関連する哲学的見方に大きな違いがあり (Gal，2019) 、「西洋」の哲学的価値観と見方も時間とともに大きく変化している (Russell,

[20]国際的なサイバーセキュリティ実行と準則に合意がないと、人工知能の影響を強く受けるデジタルセキュリティ分野(Brundage et al.， 2018)も大きな課題に直面するかもしれない (国連総会， 2015)。

1945)。総じて言えば、「世界の価値観調査」[21]や「アジアバロメーター調査」[22]などのプロジェクトで確認されているように、双方の倫理的および文化的価値観は常に進化している。

異なる地域における倫理的文化的伝統の違いは異なるガバナンスアプローチの基礎を築き上げると考えられている。たとえば、プライバシーに関する問題は東西の価値観における大きな違いと見なされる。そのため、米国やヨーロッパと比べて、中国はデータのプライバシーを比較的緩やかに制御していると考えられる。しかし、これらの主張は非常に曖昧であり、十分な議論や実証分析が行われていない (Ess，2005；Lü Yao-Huai，2005) ため、双方間に大きな誤解が生じている。まず、プライバシーの概念(Szeghalmi，2015)と関連する規制(McCallister，2018)について、米国とヨーロッパの間には大きな違いがある。しかし、中国の西洋社会に対する理解ではこれらの内部の違いを無視することが多く、米国の特徴に過度に注意を払っている[23]。第二に、中国のデータプライバシーに対する西洋の認識は時代遅れである可能性がある。Lv Yaohuai(2005)は、中国の情報倫理に関する文献は米国ほど成熟していないが、欧米の学者から強い影響を受けており、その進展も日々変化していると指摘している。多くの中国の学者や政策立案者は、関連する学術論文やレポートを公開し、AI 倫理とガバナンスの開発におけるデータプライバシーの重要性を強調している（北京知源人工知能研究所，2019；Fu，2019；Zeng, Lu, Huangfu，2018；Ding，2018b)。中国の関連する規制措置は、個人データのプライバシー保護の原則を提案し始めた。中国政府は、個人データのプライバシー基準に違反する 100 のアプリケーションを禁止し、数十のアプリケーションに対し、データの収集と保存方法を調整するよう修正を要求する[24]。これは、これらの国においてデータプライバシーに関する価値観・規範・規制に大きな違いがないわけではなく、このような違いに対する私たちの理解が一般的すぎて、理解が不十分であることを示している。

認識の差異は、中国の社会信用システム(SCS)の理解によっても反映されている。西洋メディア・政策界・および学者はこのシステムに細心の注意を払い、それを中国政府によるオーウェル式の社会的統制の例(Botsman，2017；Pence，2018)として、西洋世界の文化や政府とは大きく異なる価値観を代表するものと見なしている(Clover，2016)。しかし、中国と西洋のメディアは共に、これがシステムに対する誤解であると主張している。多くの学者が指摘するように、中国の SCS は 14 億人の中国国民すべてを評価する単一の統合プラットフォームとして設計されているのではなく、さまざまな視点に基づく解釈を提供する寛容な個々のプラットフォームのウェブであり、社会的信用スコアは主に金融機関によって与えられている（ビッグデータによる総合評価ではない）(Mistreanu，2019；Sithigh 、Siems。 2019)。Song（2019）は、社会信用システム(SCS)による基準の多くは、詐欺や地方自治体の汚職などの問題に取り組むように設計されていると指摘している。Chorzempa et al.(2018)はまた、「ブラックリストや広範囲にわたる監視などの社会的信用の多くの中核的要素は、米国のような民主主義国家にすでに存在している」とも述べている。中国の社会信用システム(SCS)はますます健全になるが、現在および将来の実装には依然として注意が必要としている。ただし、SCS がどのように機能し、使用され、中国国民に影響を与えているかについて、より明確な異文化間の理解があれば、関連する倫理およびガバナンスの問題に関する対話をより建設的に進めることができるだろう。

---

[21] http://www.worldvaluessurvey.org/wvs.jsp

[22] http://www.asianbarometer.org/

[23]前述した 7 月のセミナーで、多くの中国の学者がこの誤解について言及した。

[24] 2019 年 11 月、中国の公安省は個人データのプライバシー基準を満たさない 100 のアプリケーションを禁止した。2019 年には 683 のアプリケーションを調査した（国家ネットワーク情報管理局，2019)。2019 年 11 月に、中国の産業情報技術省は 41 のアプリケーションを発表し、データ規制要件を満たすために 2019 年末までに修正することを要請した（中国産業情報技術省，2019 年)。2018 年 7 月、中国の山東省では、個人情報に違反した 11 社の訴訟を報告した (Ding，2018c)。

米国・ヨーロッパ・中国は長い間、知識や文化的背景を共有していないため、地域間で誤解があることは驚くことではない。 したがって、性急な行動による調整不可能な根本的不一致に注意すべきである。現在、双方にも誤解が存在する。たとえば、世論調査データを分析すると、中国と米国の国民はお互いの国の特性や特徴について多くの誤解があることが分かる (Johnston, Shen, 2015)。上記したように、中国では西洋社会の多様性も単一のアメリカ式の生活パターンにまで単純化されすぎることがよくある。同時に、米国とヨーロッパは長い間、中国に対する包括的な理解に欠けており（Chen, Hu, 2019）、中国の自由化（或いは自由化の欠如）または経済成長サイクルの予測を繰り返し失敗した（「エコノミスト」, 2018; Cowen, 2019; Liu, 2019)。言語の障壁は西洋諸国にとって、AI 開発・倫理およびガバナンスにおける中国の発展を把握する時の大きな障害となる (Zhang, 2017; Ding, 2019)。Andrew Ng が 2017 年の「大西洋月刊」のインタビューで指摘したように、「言語の問題により、一種の非対称性が生み出される。中国の研究者は通常英語を習得しており、すべての英語文献にアクセスできる。 一方、英語圏のコミュニティは、中国の AI コミュニティの仕事にアクセスするのは更に難しい」 (Zhang, 2017)。たとえば、Tencent がリリースした人工知能戦略に関する本（Tencent Research Institute et al., 2017）には、倫理・ガバナンス・社会的影響の詳細な分析が含まれているが、英語の報道ではほとんど言及されていない (Ding, 2018a)。AI の研究開発に対する中国の公共投資のレベルなどの経験的な問題でさえ、米国で広く報告されているデータは、桁違いに不正確である可能性がある (Acharya, Arnold, 2019)。

最近公開された「北京 AI 原則」（北京知源人工知能研究院, 2019）と世界中で推進されている他の類似した原則(Cowls, Florid, 2018)は、中核的な課題に関して多くの重複が存在する (Zeng, Lu, Huangfu, 2018; Jobin, Ienca, Vayena, 2019)。「北京 AI 原則」では他の文献に含まれる中心的な概念と価値に明確に言及している。例えば、AI は「すべての人類に利益をもたらす」べきであり、「人間のプライバシー・尊厳・自由・自律および権利」を尊重し、 そして「可能な限り公平でいて、システムにおける差別と偏見を減らし、その透明性を向上させ、その解釈可能性と予測可能性を改善すべきである。さらに、「北京 AI 原則」と「新世代 AI ガバナンス原則」はどちらも、AI 開発にはオープン性と協力が必要であると提案し、後者は特に「学際的・領域間的・地域横断的・国境を越えた交流と協力」を奨励している (Laskai, Webster, 2019)。しかし、実際には文化的背景が異なる国では、同じ原則に対しても、解釈が異なり、注意の度合いも異なる (Whittlestone et al., 2019)。これは誤解を生むより深い原因であるかもしれない。たとえば、「西洋」の文化は「東洋」の文化よりもプライバシーを重視していると単純に想定することはできない。代わりに、プライバシーはセキュリティ等のような他の重要な価値があるものと認められない場合では、異なる地域におけるその優先順位を決める方法について、より深く突っ込んで研究すべきである(Capurro, 2005)。 同様に、多くの文化では人間の自律性を重視しているが、異なる背景でこの価値観の背後に伝えられた深い意味合いや哲学的理念も探求すべきであろう (Yunping, 2002)。

国家間の誤解が定着されて久しい。より健全な異文化間協力を築き上げるためには、まず人工知能の倫理に密接に関連する誤解を認識し、紛争の焦点とガバナンス方案に最も影響を与える紛争について相互理解を深めるべきである。ここではガバナンスの違いではなく、倫理的な違いを明確すべきである。なぜなら、いくつかのケースでは、異なるグループは独自の倫理観を持っているが、特定のガバナンス基準にも同意できるからである。これについては後で説明する。これにより、AI に直接関連する誤解（他の国の技術投資、またはデータ保護法への誤解など）を、より広範にわたる社会的・政治的または哲学的な問題など間接的に関連する誤解と区別することにも役立ち、異なる問題を解決するには異なる対策を必要とするからである。

誤解の重要性を強調することは、AI 倫理とガバナンスにおける異文化間の矛盾のすべてが根本的に誤解に基づくことを示すものではない。個人・社会・国家間の関係や、市民・民間・軍事部門の統合の度合いと特徴、および社会政策に関連する様々な特定の問題について、地域間の根深い違いを変えることは困難である。ただし、最初から誤解を減らすことに焦点を当てること

は、これらの根本的な違いをより明確に特定し、同時に十分な合意を得て効果的に協力できる環境を見つけることに有益である。これは、AI の倫理とガバナンスにおける異文化協力の課題に取り組むための重要な最初のタスクである。

意見の不一致から協力する方法の構築

AI 倫理・ガバナンスおよびより広範な社会問題に関する見解の不一致は排除できないが、合意と協力はなお可能である。前述したように、AI 倫理とガバナンスが直面する主な課題は、規範・標準およびシステムに関する異文化間の合意の達成が特に重要である領域を特定する。これらの領域では、さまざまな見解と方法に対応または推奨できることが求められる。この課題自体は異文化間の協力によって解決できる。そしてこの課題は、さまざまな文化的背景における AI の影響に対する理解、および異なるグループのニーズと欲求から情報を収集する。「新世代 AI ガバナンス原則–責任ある AI の開発」では、上記の対策を詳しく説明し、「国際的な対話と協力を展開し、 AI ガバナンスに対する各国の原則と実践を完全に尊重することを前提として、幅広いコンセンサスを備えた国際的な人工知能ガバナンスの枠組みと基準の形成を促進する」(Laskai, Webster 2019)。

抽象的レベルの倫理的認識や高レベルの原則における地域的及び文化的違いは、規範とガバナンスの特定側面での合意達成を必ずしも妨げるものではない。基本的な倫理問題についてコンセンサスに達することができないと、実質的な契約に署名することができなくなり、「核兵器禁止条約」などの多くの重要な国際合意は実現できない。法学分野における「不完全に理論化された合意」(Sunstein, 1995)は、基本的または抽象的な見解が合意に達しない場合でも、人々が特定のケースを解決する場合にはその見解が一致していることがあると指摘する。これが法執行の鍵であり、広義の多元社会にとっても重要である。多くの学者は、異文化情報倫理に関する議論では「オーバーラップするコンセンサス」(Rawls, 1993)の到達を目的とした関連概念に言及する。即ち様々なグループや文化が同じく一連の基準または実用的なガイドラインをサポートする出発点は決して同じではない (Taylor, 1996; Søraker, 2006; Hongladaro, 2016)。例えば、Taylor(1996)は国際的に認められた人権規範をさまざまな文化的伝統に導入する方法を探究した。西洋の哲学は人間の主体、および宇宙における人間のユニークな位置などの本質的な問題に対し、仏教などの他の哲学システムとは見方が異なるが、どちらの哲学システムも最終的に同じ人権規範を明らかにできる。

Wong(2009) は、異文化を共有しながら共通の基盤を求める異文化情報倫理について合意に達することは非現実的であると批判し、これに従って策定された規範はあまりにも「弱く」、包括的かつ規範的な内容が不足していると考えている。Søraker(2006)は情報倫理における「実用的」方法に対し、上記のような反対の意見に言及した。即ち、このような合意は実質的かつ規範的な内容に完全には基づいていないため、より脆いものになるかもしれない。しかし、これらの異論に対する Søraker の対応からわかるように、「オーバーラップするコンセンサス」は、一致した規範と実用的なガイドラインに達することを目的とし、多くの哲学的かつ規範的な見解によって生成およびサポートされているため、より強力である。しかし、このような状況と明確に区別されるべきは、特定の文化が実際の議論を通じて自分の価値を他の文化に押し付けようとする場合である。或いは幾つかの文化グループは合意に達しているが、何らかの理由で規範的な内容が形成されていない場合もある。本論文は、後者の場合ではそのような状況が心配すべきであると考え、Wong(2009)の見方に同意する。Taylor(1996)が提案した事例では、人権が多くの哲学的見解によって支持されていると示す。これにより、「オーバーラップするコンセンサス」の合理性が証明できるものであるとする。

この論文では、人工知能が人間にとっての有益性を確保するため、共通の認識を持つ基本的な価値について国際的コンセンサスに達するよりも、規範と実践的なガイドラインにおいてオーバーラップするコンセンサスが存在する領域の探索が重要であると考えている。 これはまた、最

近多くの提案を支える目標でもある[25]。高レベルの倫理原則についてコンセンサスに達しても、これらの原則が完全に正当化されるわけではない（(Benjamin，1995)。堅固で信頼性の高い人工知能の規範・標準・および規制を確立するための一番優れた方法は、さまざまな価値観体系が尊重するコンセンサスを提案することである。

原則の着実的な実装

相互理解を促進し、理論的にガバナンスアプローチについてコンセンサスを得たところでも反対される可能性は存在する。もっとも反対意見が実際の AI の開発と応用に影響を与えることは少ないと考えている。なぜなら、そのようなコンセンサスに達するには、強力な国や企業の行動に影響を与えなければならない一方、これらの国または企業は協力する意思がないためである。

国、企業、およびその他の関係者間の複雑な権力の変化は、関連する AI 倫理とガバナンスの問題にも密接に関連している。しかし、この論文では、この障壁が私たちの提案を損なうと考えられない理由を簡単に説明するのみとし、詳述しない(ただし、この現象をさらに調査する価値がある）。歴史的な先例によると、公的および学術的グループの影響力を利用し、重要かつグローバルな問題を解決するため権力者を促すことは難しいが、それは実現可能である。実証的研究は、広範囲にわたる異文化間における「専門家グループ」（特定の分野の専門家ネットワークなど）が国際政策における効果的な協力を促進できることを示している(Haas，1992)。たとえば、武器管理専門家グループは、冷戦中の核兵器管理に関する国際的なコンセンサスを促進し、米国が旧ソビエト連邦との協力関係を確立することを支援した(Adler，1992)。そして生態学の専門家グループは国家政策の調整に成功し、成層圏のオゾン層を保護した(Haas，1992)。

人工知能の分野では、会社員の積極的な行動及び国際的な学術研究と活動はすでに大手企業や国々の取り組みに影響を与えており、特に人工知能の軍事分野における応用の場合、その影響が最も大きい。国際学会や民間コミュニティの専門家は、国際ロボット兵器管理委員会（ICRAC）や致命的なロボット工学の禁止など、多くの大規模なキャンペーンを開始し、戦争における人工知能の適用に関する懸念を表明した。これらの運動は、国連通常兵器条約会議（クラスター爆弾、レーザーブラインド兵器、および地雷の禁止について交渉を行う会議であり、CCW とも呼ばれる）での致命的な自律兵器(LAWs)の議論を推し進めてきた (Belfield，2020)。90 か国が致命的な自律兵器に対する立場（ほとんどは国連通常兵器条約会議で提案されたもの）を表明し、28 か国が致命的な自律兵器の使用を禁止することに合意した[26]。2018 年には、4,000 人を超える Google の従業員が抗議請願書に署名し、映像分析における AI の使用を模索する軍事プロジェクトと呼ばれるペンタゴン Maven 人工知能プロジェクトへの Google の参加に抗議し、辞任する従業員もいた(Conger，2018)。米国・ヨーロッパ・日本・中国・韓国などの学者も活動グループを作り、記事や公開書簡を公開し、Google 請願の従業員を支援した(ICRAC 2018)[27]。Google はすぐに Maven プロジェクトとの契約更新終了を発表し、米国国防総省の 100 億ドルのクラウドコンピューティング契約の入札からも撤回する(Belfield，2020)。

より広範なレベルでは、国際的な学術コミュニティと民間グループの努力が原則の規制に貢献でき、将来、より拘束力のある規制の開発に強固な基盤を提供した。たとえば、欧州委員会は

---

[25]例えば、Florid らは 2018 年に「統一されたフレームワーク」を提案した。Awad らは 2018 年に「機械倫理の世界的および社会的受容のための原則」について議論した。Jobin、Ienca と Vayena は 2019 年に倫理的人工知能の構築に関する「グローバル合意」を調査した。

[26]中国は、戦場での完全自律兵器の使用禁止を支持しているが、完全自律兵器の開発に反対しているわけではない。 米国、ロシア、英国、イスラエル、フランスはこの禁止に反対した(Kania，2018)。

[27]国際ロボット兵器管理委員会は、何千人もの学者や研究者によって署名された公開書簡を公開した。致命的なロボットの禁止運動のメンバーはまた、Google 請願従業員をサポートするため、公開記事や公開書簡によって企業のリーダーに連絡した。https://www.stopkillerrobots.org/2019/01/rise-of-the-tech-workers/ を参照されたい。

「人工知能白書−卓越性と信頼を求めるヨーロッパの方法」（欧州委員会，2020 年）を発行し、「将来の EU 規制フレームワークのポリシーの選択により、関係者に適用される法的要件のタイプを決める」と提案し、その中リスクの高いアプリケーションを特に強調する（欧州委員会，2020b）。この文書は、AI に関する EU の高レベル専門家グループ（学界、産業界や民間グループの 52 人のヨーロッパ専門家によって構成）の活動に強く影響を受けた[28]。同様に、米国国防総省は、米国の学術・産業界・政府の利害関係者と 15 か月議論した結果、AI に関する倫理原則（米国国防総省，2020）を正式に採用し、具体的な仕事を遂行するために担当者を指定した(Barnett 2020)。上記のどちらのケースも、相談したグループは特定の地域に限定されているが、異なる地域で開発された原則の整合性と重複は、幅広い知識グループとの地域間の交流が本質的に多くのアイデアと提案を提供できることを示している。これは、異文化間の協力とコンセンサスから得た洞察が、地域レベルおよび国レベルで規制の枠組みに組み込まれる可能性があることを示している。

5. 推奨事項

アカデミアは、AI 倫理とガバナンスに関する異文化間の協力をサポートする上で重要な役割を果たしている。学術研究によって、協力が最も必要な分野と協力のタイプを探し、プログラムを策定し、協力におけるより現実的な障壁を乗り越えることができる。本論文では、さまざまな学術的専門知識を必要とする多くの質問を提起し、次のものが含まれる。何が地域間の協力を妨げる最大の誤解であるのか。AI 倫理とガバナンスに基づく国際協定はどこに必要なのか。 倫理的な問題の違いにもかかわらず、特定のガバナンス基準に対しどのように合意に達するか。学術コミュニティは、自由に流れる国際的及び異文化間のアイデア交換の伝統により、実際に地域間と文化間の相互理解を深めることに特に適している。学者は、業界や政府で働く人々にとっては常に困難な方法でも、国際的な同僚とオープンな会話ができる。世界中からの 2 人の学者は、他の国の政府や企業に対して強い批判があっても、実りある協力ができる。

以下の推奨事項は、AI 倫理とガバナンスに関する異文化間の理解と協力を促進するために、学術センター・研究機関・および研究者たちが実行できる幾つかのステップを示している。 これらの分野では、既にいくつかの優れたプロジェクトが順調に進んでいる。ただし、この論文では、AI の新しい分野や地域に応用されるペースに注目する価値があると信じており、学術コミュニティに、より広い倫理とガバナンスの研究プロジェクトにおいて異文化間の橋を構築し、さらに異文化の専門知識を組み込むよう呼びかけている。

異文化協力に基づく AI 倫理とガバナンスに関する研究課題を策定する。異文化共同研究プロジェクトの推進は、国際的な政策協力を支える国際的な研究コミュニティの構築に重要な役割を果たすであろう(Haas，1992)。

このような協力の要件を満たす研究プロジェクトは、対照的かつ将来を見据えた活動を実行し、異なる文化における AI の社会的影響に関する前向きなビジョンと明確な懸念の相違点を探るべきである。これは、国際的なビジョンを開発するのに役立ち、AI 開発で達成または回避すべき側面を明確にし、倫理とガバナンスのフレームワークに関するより実践的な議論を導くことができる。コンセンサスの達成は大きく影響しているかもしれない。安全性とセキュリティは世界中の人間の文化の基本であり、文化的脅威を回避するため合意を築くことは、より簡単な出発点になる。しかし、本稿では前向きなビジョンを重視することも重要であると考えている。 異文化間の学者たちは人間が共有するより良い未来を創造することは、共通の価値観におけるニュアンスを掘り下げる良い方法であろう。

[28] 展開およびメンバー情報はこのリンクを参照。https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

また、途上国に対する AI の継続的かつ予想される影響を、これらの国の専門家と協力して検討することも非常に重要である。そのような研究は、地元の専門家の指導の下で発展途上国における技術展開の意思決定を行うべきである。こうして、意思決定の力は地元のグループの手に委ねることができる (Hagerty 、Rubinov，2019)。より実用的なレベルからいうと、国際的な研究機関は効率的な協力を通じて、AI の安全性・セキュリティおよび社会的危害への回避に関する研究・専門知識・およびデータセットが含まれる国際的な共有フレームワークを構築する。

異文化協力そのものをさらに推進するためには、異なる地域や文化団体間の研究者の協力も不可欠である。本文の議論（特にセクション 4）では、さらに調査する必要のある多くの研究領域を指摘する。それは以下の通り：

- 異なる文化における人工知能倫理とガバナンスの価値観・仮説・優先順位に関する一般化した違いを調査・特定し、変更する。具体的には[29]：
    - 異なる哲学的伝統に由来する技術倫理の類似点と相違点を分析し、人工知能の開発・適用・影響力および実際的なガバナンスに対する影響を調査する；
    - 様々な文化における核心的な価値観差異の背後にある経験的証拠を探る。たとえば、一部のプロジェクトでは、データのプライバシー・国家と個人の役割および技術の進歩に対する態度など、AI ガバナンスに関する東洋文化と西洋文化の間における既定の価値観の違いを特定して調査する；
    - 異なる社会における AI 応用の優先順位と制限の地域差、およびこれらの違いが AI の研究開発に与える影響を理解する。
- 綿密な分析により、AI ガバナンスがグローバルなコンセンサスを必要とする領域を特定し、異文化間の違いを許容もしくは推奨する領域と区別する；
- 異文化レベルで、国際的および世界的に AI 基準が必要とされる主要な分野で人工知能基準の策定を支援する；地域特性が国際的基準に一致することを確保するため、柔軟なガバナンスモデルを模索する；
- パタンとアプローチを調査し、基本的または抽象的な倫理的問題の違いに対応し、具体的なケース・意思決定およびガバナンス基準に対してコンセンサスに達する。他の分野で成功した事例から学び、AI 倫理とガバナンスに適用する。

これらの分野では、既にいくつかの優れたプロジェクトが順調に進んでいる。ただし、この論文では、AI の新しい分野や地域に応用されるペースに注目する価値があると信じており、学術コミュニティに、より広い倫理とガバナンスの研究プロジェクトにおいて異文化間の橋を構築し、さらに異文化の専門知識を組み込むよう呼びかけている[30]。

主要論文とレポートを翻訳する。他の科学分野と同様に、言語はまた、AI の開発とガバナンスおよび倫理に対する異文化理解を妨げる主要な現実的障壁である(Amano et al.，2016)。したがって、AI 倫理とガバナンス、および AI 分野において急成長している研究文献を多言語版に翻訳されれば、非常に価値があるであろう。AI 分野で主要なアジアの学者の多くは英語に堪能であるが、多くはそうではない。また中国語や日本語に堪能な西洋の研究者は遥かに少ない。

さらに、上記した誤解のいくつかは、他の地域の学者がある地域の主要な文献を理解し引用する方法にも密接に関連している。欧米のメディアは、2017 年に発表された中国の「新世代 AI 開発計画」を、経済的および戦略的に AI に対する世界的な主導的地位を強化する手段として説明

[29]異文化情報倫理の分野では、既に Capurro(2005，2008)、Ess (2006)と Hongladarom et al. (2009)などの、この問題に関するいくつかの優れた研究がある。ただし、この論文では、これらのトピックを中心に異文化間の研究協力が得られ、人工知能倫理に関する実践的な議論と行動に多く注意することを期待している。

[30]本論文の何人かの執筆者は、このようなイギリスと中国の異文化研究をサポートすることを目的とした取り組み計画にも参加している。https://ai-ethics-and-governance.institute/%e5%85%b3%e4%ba%8e/を参照されたい。

した(Knight, 2017；Demchack, 2019)。しかし中国にとって、AI 開発の目標は主に中国の経済と社会発展のニーズによるものであり（中国国務院，2017 年）、必ずしも国際的な競争優位ではない (Ying，2019)。主要な用語やポイントの不適切な翻訳は誤解を引き起こす。たとえば、「計画」の中国語テキストには、「中国は 2030 年までに世界の主要な AI イノベーションセンターになることを目指している」 (Ying，2019)[31] と書いてあるが、一部の英語の翻訳はこの文を中国が間もなく「世界をリードする AI イノベーションセンター」になると翻訳した(Webster et al.，2017 など)。これはさらに、Google の親会社である Alphabet の元会長 Eric Schmidt によって、「2030 年までに中国は AI 業界を支配するようになる。本当？中国政府がこう述べている」と解釈され、発表された(Shead, 2017)。この文の表現はある意味で大きな変化はないが、重要な意味合いを持っている。元の表現では、国際的な覇権とは対照的に、中国のリーダーシップと開発進展をより優しく表現している。重要な文書の高品質な多言語翻訳によれば、学者はこれらの言い換えで失われる可能性のある言語と文脈のニュアンスを見分けることができる。

文献やレポートの高品質な多言語翻訳の提供も、異文化交流を尊重し、参加する意欲を反映しており、さらなる協力を促進できるかもしれない。学術資料や政策資料の高品質な翻訳は複雑で時間のかかる作業であるが、より強くサポートされ、承認されると期待している。今は賞賛すべき仕事が多くあり、そして勢いよく成長している。たとえば、Jeff Ding は多くの中国の主要な AI 文献(Ding，2019)を翻訳した。中国海国図知研究院の国際関係・科学技術・その他のトピックについての本は、5 つの言語で出版されている(http://www.intellisia.org/)。Brian Tse は英語版の OpenAI の組織プログラムなどを中国語に翻訳した[32]。New America は中国産業情報技術省が提案した「3 年間の行動計画」を英語に翻訳した (Triolo et a.，2018)。

主要な AI に関する検討会議や倫理とガバナンス会議を異なる大陸で順次開催できるようにする。AI の開発・倫理・およびガバナンスへのグローバルな参加を促進するため、これらのトピックに関する主要な会議やフォーラムを異なる大陸で交互に開催することを提案した。 これにはいくつかの利点がある。第一は一部の地域の学者が他の場所へ会議に急いで参加するために常に多くの時間とお金がかかることを避けることができる。第二は異なる地域のグローバルな研究グループに様々な程度で影響を与えるビザ制限を避けることができる。第三は海外に出たくない地元の研究グループを引きつけ、地元の主催者が積極的に参加するよう推奨する。第四は主催者はイベントを単一言語ではなく多言語で開催することを推奨する。

他の積極的な対策もある。AI 研究会議のうち、国際 AI 共同会議(IJCAI)は 2019 年にマカオで、2013 年に北京で開催された。これらは中国で開催された最初の 2 つの国際 AI 共同会議である（日本では 2 回開催されていた）。国際機械学習会議（ICML）は 2014 年に北京で開催され、2021 年にソウルで開催される予定である。国際表現学習会議（ICLR）は 2020 年にエチオピアで開催される予定で、アフリカにとって最初の重要な機械学習会議である。AI 倫理とガバナンスは比較的新しい分野であるため、そのようなトピックをめぐって明示的に開催される大規模な会議は多くないが、もし可能であれば、これらの会議が順番に異なる大陸で開催され、グローバルな参加を確保することが特に重要である。例えば、人工知能と倫理および社会に関する会議は、人工知能促進会（AAAI とも呼ばれ、全身は米国人工知能学会）によって主催されるため、現在米国で開催される。 これらのトピックに対し国際的な学術コミュニティを構築するには、このような状況を変えなければならない。北京知源大会（北京 AI カンファレンス学会シリーズ）をはじめ、中国では AI の倫理とガバナンスに焦点を当てた会議が急速に発展している。AI の倫理とガバナンスに関連するトピックをカバーする既存の会議も幾つかあり（ただし、それらについては明示されていない）、国際的な参加をさらに強化することができる。情報社会世界サミットフォーラム（主にジュネーブで開催）、インターネットガバナンスフォーラム(Internet Governance

[31] 2019 年 7 月のセミナーでは、一部の参加者もこの翻訳を提供した。
[32] https://openai.com/charter/

Forum)、権利大会(RightsCon)などが挙げられる（後者の 2 つは南アメリカ、インド、アフリカを含む非常に多くの場所で開催されたが、 東アジアでは開催されていない）[33]。

博士課程の学生とポスドクのための共同または交換プログラムを確立する。異文化背景を持つ研究者にキャリアの早い段階で異文化協力への参加を推奨することは、協力と相互理解を強化し、研究進捗を推し進めるのに役立つ。今、多くの国際的な奨学金と交換プログラムがさまざまな国の間に確立され、最も一般的に見られるのは中国と米国の間のプロジェクトである（たとえば、米国と中国の「知行中国」奨学金プログラムと蘇世民奨学金プログラム）。他に、イギリスとシンガポールの間にもプロジェクトがある（たとえば、キングスカレッジロンドンとシンガポール国立大学が設立した哲学または英語の共同博士号プログラム）。報道によると、これらの計画は特に人工知能を目的としていないそうである。 現在、知られている AI 倫理とガバナンスのプロジェクトは、博古睿研究院の学者計画と「世界中の学者」国際的奨学金プログラムだけである(Bauch，2019)[34]。そのようなプログラムをさらに確立することは、AI の将来の国際協力を促進することができ、今は学ぶための多くの既存のモデルとイニシアチブがある。

本文では広い視点から、人工知能と機械学習における学際的な専門家の国際交流とコラボレーションをサポートするため、特定のビザ処理チャネルの確立、ビザプロセスの簡素化と迅速化、およびプロセスの公正かつ標準的であることの確保など、人工知能協力組織による関連提案、そして政府への提言をサポートしている。これらの推奨事項には、AI 倫理とガバナンスに携わっている、または携わる予定の専門家が含まれる（「技術的作業」のカテゴリに含まれない場合もる）(PAI Staff， 2019)。

6. 制限と将来の方向性

AI 倫理とガバナンスにおける異文化協力の促進については学界が重要な役割を果たすと考えている。相互理解と協力に基づくコミュニティは、基本的な価値の違いをすべて排除することなく効果的に確立できる。そして文化団体間の誤解を減らすことは特に重要であるかもしれない。この論文の提案では、異文化協力における多くの障壁を乗り越えることはできず、AI が全世界にとって有益であることを確保するために、さらに多くの作業を必要していることを心得ている。この目標を導く 2 つのより広い将来の研究の方向性を簡単に述べる。

異文化協力の障壁、特に権力のダイナミクスと政治的緊張に関する内容をより注意深く分析する。歴史的な成功事例の分析は、AI 倫理とガバナンスに基づく異文化活動が、実際のガイドライン・標準・規制の策定に大きな影響を与える可能性があることを示すが、実装と実施のレベルでは多くの本稿では言及できない障害が存在する。将来の研究では、歴史を振り返り、過去のグローバルな規範や制度に影響を与えたイベントがいつ、どのように発生したかを調査すべきである。そのような研究は非常に価値があると考えている。

上記で指摘したように、異文化間における価値観の違いや誤解に加え、権力関係や政治的緊張に関連するさまざまな問題が、異文化協力に大きな障害を引き起こす可能性がある。これらの問題が AI の倫理とガバナンスにおける異文化協力をどのように阻害するかについての研究は、協力を進める上での学術プロジェクトの限界を認識し、これらの方法が権力および政治上の変動分析との融合を理解するのに役立つため、非常に価値がある。

[33]産業クラスターは、国際的なセミナーや会議への参加を通じて異文化協力を促進する。この論文では、人工知能産業開発同盟の進捗状況をこの分野における最近の取り組みの例として参照することを勧める。http://www.aiiaorg.cn/を参照されたい。

[34]長期的優先事項センターによって設立された「世界中の学者」プログラムには、人工知能のセキュリティとガバナンスのトピックも含まれている。

将来の強い AI システムが異文化協力において直面する課題を検討する。将来の AI の発展は、グローバルな協力に新たな大きな課題をもたらすかもしれない。一部の学者は、AI の将来の開発が産業革命や農業革命のように変革的な影響をもたらす可能性があることを示唆している。(Karnofsky，2016；Zhang, Dafoe，2019)。厳密なグローバルな方向制御がなければ、この技術の進歩は、技術的に進歩した国と遅れている国との間で富と権力の前例のないギャップにつながるかもしれない。他の学者の研究はより未来志向であり、超人工知能を備えたシステムを開発する可能性を提案している（即ち人間の知能を超えた一般的な知能；Bostrom， 2014）。結果と安全性を考慮しなければ、そのようなシステムの強い機能は、人類の文明に壊滅的なリスクをもたらす恐れがある。一部の学者は、破滅的な結果を回避するための重要な対策として、統一された価値を達成し、人間の価値観に適合するシステムを設計することを示唆している (Russell，2019)。共通の価値観と原則についてグローバルな合意に達し、複数の価値観が尊重されるシステムを設計することが最優先事項となっている。

多くの専門家はこの技術の将来の開発に対して、かなり異なる見方を持っていて、ほとんどの人はそれが数十年かかると信じている。しかし、効果的なコラボレーション結果を達成するには、協力関係を築き、必要な合意に達するまでに数十年の努力が必要であるかもしれない。これは、今の協力計画が、目の前の AI システムの倫理およびガバナンスにおける課題を解決するだけでなく、将来の課題を予測して対応するための基礎を築く必要があることを示している。

7. まとめ

グローバル社会において AI の最大のメリットを実現するには、学際的・国境を越えた異文化間にわたるより深い協力次第である。現在米国・ヨーロッパ・中国の間での緊張と信頼感の欠如は、特に協力を制限している。誤解がこの不信感を悪化させており、社会的政治的優先順位の違いが過度に強調されているか、誤解されている。さらに、これらの地域が AI に関連するすべての重要な倫理原則に同意し、そしてルールと標準に組み込むことができるという考えは、あまりにも単純すぎている。これにしても、これらの地域がグローバルな AI 倫理とガバナンスを形成する上で過度に支配的であることは望ましくない。AI の影響を受ける全世界のすべてのコミュニティが含まれ、権限を与えられている必要がある。「AI の超大国」間の相互理解を深めるための取り組みは、2 つの利点があると考えている。1 つ目は、グローバルな AI ガバナンスに関する主要な緊張を緩和すること。２つ目は、倫理とガバナンスのフレームワークの開発を支援する参考経験を提供することよって、複数の価値観をサポートし、適切なコンセンサスを達成すること。うまく機能している AI のグローバルな連携システムでは、課題は新しいモデルを開発することにある。即ち国際的なコンセンサスによって構築およびサポートされている原則と標準だけではなく、様々な社会的ニーズを満たすために研究と政策策定のコミュニティによって提案された色々な方法も含まれるべきであろう。

より実用的な視点から、国際 AI 研究コミュニティ、AI 倫理およびガバナンスコミュニティは、自分たちの活動がどのようにグローバルな協力をサポートできるか、地域間のさまざまな社会的見解とニーズに対する理解をどのように促進するかについて慎重に検討しなければならない。広範囲にわたる異文化間の研究協力と交流、異なる地域で開催される会議、および多言語の出版物は、協力の障壁を減らし、異なる視点と共通の目標を理解する障壁を取り除くことに役立つであろう。政治的傾向が孤立主義に偏っていく中、研究者たちが世界的に有益な AI を目指して、国や文化の格差を越えて取り組む努力がこれまでになく重要であると考えられている。

謝辞

本論文の完成に関し、筆者は 2019 年 7 月 11 日から 12 日まで行われた異文化セミナー「人工知能のための有益な信頼メカニズムの構築」の参加者全員に感謝の意を表する。彼らは会議で本

論文に関連する多くの重要な議論を提供してくれた。本論文の最初の草稿に対して、有益なコメントを提供してくれた Emma Bates、Haydn Belfield、Martina Kunz、Amritha Jayanti、Luke Kemp、Onora O'Neill および 2 人の匿名査読者にも感謝の意を表する。

参考文献

Acharya, A. (2019). Why International Ethics Will Survive the Crisis of the Liberal International Order. *SAIS Review of International Affairs*, 39(1), 5-20.

Acharya, A., & Arnold, Z. (2019). Chinese Public AI R&D Spending: Provisional Findings. Centre for Security and Emerging Technologies Issue Brief. Available at: https://cset.georgetown.edu/wp-content/uploads/Chinese-Public-AI-RD-Spending-Provisional-Findings-2.pdf Accessed 23 December 2019

Adler, E. (1992). The emergence of cooperation: national epistemic communities and the international evolution of the idea of nuclear arms control. International organization, 46(1), 101-145.

Allen, J. R., Husain, A. (2017). The Next Space Race is Artificial Intelligence. *Foreign Policy*. Available at: https://foreignpolicy.com/2017/11/03/the-next-space-race-is-artificial-intelligence-and-america-is-losing-to-china/ Accessed 21 December 2019.

Altmann, J. (2019). Autonomous Weapon Systems–Dangers and Need for an International Prohibition. *In Joint German/Austrian Conference on Artificial Intelligence* (Künstliche Intelligenz) (pp. 1-17). Springer, Cham.

Amano, T., González-Varo, J. P., & Sutherland, W. J. (2016). Languages are still a major barrier to global science. *PLoS biology*, 14(12), e2000933.

Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94(886), 687-709.

Askell, A., Brundage, M., & Hadfield, G. (2019). The Role of Cooperation in Responsible AI Development. arXiv preprint arXiv:1907.04534.

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.F. & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59.2018

Barnett, J. (2020). DOD hires policy team to implement AI principles. Available at: https://www.fedscoop.com/dod-hires-new-ai-policy-team/ Accessed March 12 2020.

Bauch, R. (2019). Berggruen Institute announces 2019-2020 class of Fellows in U.S. and China as international cohort of Berggruen Thinkers to Study Great Transformations. Berggruen Institute. Available at: https://www.berggruen.org/news/berggruen-institute-announces-2019-2020-class-of-fellows-in-u-s-and-china-as-international-cohort-of-berggruen-thinkers-to-study-great-transformations/ Accessed 27 December 2019.

Beijing Academy of Artificial Intelligence. (2019). Beijing AI Principles. Available at: https://www.baai.ac.cn/blog/beijing-ai-principles Accessed 24 December 2019

Belfield, H. (2020). Activism by the AI Community: Analysing Recent Achievements and Future Prospects. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (pp. 15-21).

Benaich, N., and Hogarth, I. (2019). State of AI Report 2019. Available at https://www.stateof.ai/. Accessed 19 December 2019

Benjamin, M. (1995). The value of consensus. *Society's Choices: Social and Ethical Decision Making in Biomedicine*. National Academy Press

Bostrom, N. (2014) *Superintelligence: Paths, Dangers, Strategies* (Oxford Univ. Press).

Botsman, R. (2017). Big data meets Big Brother as China moves to rate its citizens. *Wired UK*, 21.

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Campaign to Stop Killer Robots (2018). Country Views on Killer Robots. Available at: https://www.stopkillerrobots.org/wp-content/uploads/2018/11/KRC_CountryViews22Nov2018.pdf Accessed 11 March 2020.

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). AI Now 2017 report. AI Now Institute at New York University. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf Accessed 18 December 2019

Capurro, R. (2005). Privacy. An intercultural perspective. Ethics and information technology, 7(1), 37-47.

Capurro, R. (2008). Intercultural information ethics. The handbook of information and computer ethics, 639.

Cave, S., & Ó hÉigeartaigh, S. (2018). An AI race for strategic advantage: rhetoric and risks. *Proceedings of the 2018 AAAI/ACM Conference on Artificial Intelligence, Ethics and Society.*

China State Council (2017). New Generation Artificial Intelligence Development plan. Available at: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm (translation: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/). Both accessed 20 December 2019.

Chen, D., & Hu, J. (2019) No, There Is No US-China 'Clash of Civilizations' The Diplomat. Available at: https://thediplomat.com/2019/05/no-there-is-no-us-china-clash-of-civilizations/ Accessed December 2016.

Chorzempa, M., Triolo, P., & Sacks, S. (2018). China's social credit system: A mark of progress or a threat to privacy? *Policy Briefs PB18-14*, Peterson Institute for International Economics.

Cihon, P. (2019). Standards for AI Governance: International Standards to Enable Global Coordination in AI Research & Development. *Future of Humanity Institute Technical report*. Available at: https://www.fhi.ox.ac.uk/wp-content/uploads/Standards_-FHI-Technical-Report.pdf. Accessed 20 December 2019.

Clover, C. (2019). China: When big data meets big brother. *Financial Times*. Available at: https://www.ft.com/content/b5b13a5e-b847-11e5-b151-8e15c9a029fb.

Conger, K. (2018). Google employees resign in protest against Pentagon contract. Available at: https://gizmodo.com/google-employees-resign-in-protest-against-pentagon-con-1825729300 Accessed 11 March 2020

Cowen, T. (2019). What If Everyone's Wrong About China? *Bloomberg*. Available at: https://www.bloomberg.com/opinion/articles/2019-08-19/china-s-liberalization-shouldn-t-be-ruled-out-just-yet

Cowls, J., & Floridi, L. (2018). Prolegomena to a White Paper on an Ethical Framework for a Good AI Society. SSRN preprint.

Demchak, C. C. (2019). China: Determined to dominate cyberspace and AI. *Bulletin of the Atomic Scientists*, 75(3), 99-104.

Ding, J. (2018a). ChinAI #1 Available at: https://mailchi.mp/b945e27a35ff/chinai-newsletter-1-welcome Accessed 30 December 2019

Ding, J. (2018b). Deciphering China's AI dream. *Future of Humanity Institute Technical Report*. Available at: https://www.fhi.ox.ac.uk/wp-content/uploads/Deciphering_Chinas_AI-Dream.pdf Accessed 19 December 2019

Ding, J. (2018c). ChinaAI #19: Is the Wild East of big data coming to an end? A turning point case in personal information protection. ChinAI Newsletter. Available at: https://chinai.substack.com/p/chinai-newsletter-19-is-the-wild-east-of-big-data-coming-to-an-end-a-turning-point-case-in-personal-information-protection Accessed 28 December 2019

Ding, J. (2019). ChinAI #48: Year 1 of ChinAI. ChinAI Newsletter. Available at: https://chinai.substack.com/p/chinai-48-year-1-of-chinai Accessed 26 December 2019

Ess, C. (2005). Lost in translation?: Intercultural dialogues on privacy and information ethics." *Ethics and Information Technology* 1: 1-6.

Ess, C. (2006). Ethical pluralism and global information ethics. *Ethics and Information Technology*, 8(4): 215–226.

European Commission (2020). On Artificial Intelligence - A European Approach to Excellence and Trust. White Paper. Available at: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf Accessed March 11 2020.

European Commission (2020b). https://ec.europa.eu/digital-single-market/en/artificial-intelligence Accessed March 11 2020.

Feldman, M. D., Zhang, J., & Cummings, S. R. (1999). Chinese and US internists adhere to different ethical standards. *Journal of General Internal Medicine*, 14(8), 469-473.

Gal, D. (2019). Perspectives and Approaches in AI Ethics: East Asia. *Oxford Handbook of Ethics of Artificial Intelligence*, Oxford University Press, Forthcoming.

Gehrke, J. (2019). State Department preparing for clash of civilizations with China. *The Washington Examiner*. Available at: https://www.washingtonexaminer.com/policy/defense-national-security/state-department-preparing-for-clash-of-civilizations-with-china Accessed 22 December 2019.

Gries, P. H. (2009). Problems of misperception in US-China relations. *Orbis*, 53(2), 220-232.

Haas, P. M. (1992). Introduction: epistemic communities and international policy coordination. *International Organization*, 46(1), 1-35.

Hagerty, A., & Rubinov, I. (2019). Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence. arXiv preprint arXiv:1907.07892.

Haynes, A. & Gbedemah, L. (2019). The Global AI Index: Methodology. Available at: https://www.tortoisemedia.com/intelligence/ai. Accessed 21 December 2019

Hongladarom, S., Britz, J., Capurro, R., Hausmanninger, T., & Nakada, M. (2009). Intercultural information ethics. International Review of Information Ethics, 11(10), 2-5.

Hongladarom, S. (2016). Intercultural information ethics: A pragmatic consideration. In Information Cultures in the Digital Age (pp. 191-206). Springer VS, Wiesbaden.

Houser, K. (2018). US military declares mandate on AI. Futurism. Available at: https://futurism.com/the-byte/jaic-militarys-ai-center. Accessed 22 December 2019

Hudson, R. (2019) France and Canada move forward with plans for global AI expert council. *Science Business*. Available at: https://sciencebusiness.net/news/france-and-canada-move-forward-plans-global-ai-expert-council. Accessed 27 December 2017

International Committee for Robot Arms Control (2018). Open Letter in Support of Google Employees and Tech Workers. Available at: https://www.icrac.net/open-letter-in-support-of-google-employees-and-tech-workers/ Accessed 11 March 2020.

Jervis, R. (2017). *Perception and Misperception in International Politics: New Edition*. Princeton University Press.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.

Johnston, A. I., & Shen, M. (Eds.). (2015). Perception and misperception in American and Chinese views of the other (p. 63). Washington, DC: Carnegie Endowment for International Peace.

Jun, Z. (2018). The West exaggerates China's technological progress. *Nikkei Asian Review*. Available at: https://asia.nikkei.com/Opinion/The-West-exaggerates-China-s-technological-progress Accessed 30 December 2019.

Kania, E. (2018). China's Strategic Ambiguity and Shifting Approach to Lethal Autonomous Weapons Systems. *Lawfare, April*, *20*.

Karnofsky, H. 2016. Potential Risks from Advanced Artificial Intelligence: The Philanthropic Opportunity. Available at: https://www.openphilanthropy.org/blog/potential-risks-advanced-artificialintelligence-philanthropic-opportunity. Accessed 9 March 2020

Knight, W. (2017). China plans to use artificial intelligence to gain global economic dominance by 2030. *MIT Technology Review*. Available at: https://www.technologyreview.com/s/608324/china-plans-to-use-artificial-intelligence-to-gain-global-economic-dominance-by-2030/ Accessed 26 December 2019

Laskai, L. & Webster, G. (2019). Translation: Chinese Expert Group Offers 'Governance Principles' for 'Responsible AI'. *New America, DigiChina*. Available at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/translation-chinese-expert-group-offers-governance-principles-responsible-ai/ Accessed 30 December 2019

Lee, K. F. (2017). The real threat of artificial intelligence. *The New York Times*, 24. Available here: https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html Accessed 30 December 2019

Liu, M. (2019). 30 Years After Tiananmen: How the West Still Gets China Wrong. *Foreign Policy*. Available at: https://foreignpolicy.com/2019/06/04/30-years-after-tiananmen-how-the-west-still-gets-china-wrong/ Accessed 19 December 2019

May, T. (2018). Transcript of keynote speech at 2018 World Economic Forum. Available at: https://www.weforum.org/agenda/2018/01/theresa-may-davos-address/. Accessed 27 December 2019

Matsakis, L. (2019). How the West Got China's Social Credit System Wrong. *Wired*. Available at: https://www.wired.com/story/china-social-credit-score-system. Accessed 19 December 2019

McCallister, J., Zanfir-Fortuna, G., & Mitchell, J. (2018). Getting ready for the EU's stringent data privacy rule. *Journal of Accountancy*, 225(1), 36-41.

McDonald, H. (2019). Ex-Google worker fears 'killer robots' could cause mass atrocities. The Guardian. Available at: https://www.theguardian.com/technology/2019/sep/15/ex-google-worker-fears-killer-robots-cause-mass-atrocities Accessed 11 March 2020.

Ministry of Industry and Information Technology of People's Republic of China. (2019). APP (first batch) notification on infringement of user rights. Available at: http://www.miit.gov.cn/n1146290/n1146402/n1146440/c7575066/content.html Accessed 29 December 2019

Mistreanu, S. (2019). Fears about China's social-credit system are probably overblown, but it will still be chilling. *Washington Post*. Available at: https://www.washingtonpost.com/opinions/2019/03/08/fears-about-chinas-social-credit-system-are-probably-overblown-it-will-still-be-chilling/ Accessed 20 December 2019

Mozur, P. (2019). One Month, 500,000 Face Scans: How China Is Using AI to Profile a Minority. *The New York Times*. Available at: https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html Accessed 20 December 2019

National Cyber Security Advisory Centre. (2019). Available at: https://mp.weixin.qq.com/s/smT4RbHsA_x0vIZjEKV_yg? Accessed29 December 2019

Ochigame, R. (2019). The invention of "ethical AI": How Big Tech manipulates academia to avoid regulation. *The Intercept*. Available at: https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/ Accessed 22 December 2019

Oppenheimer, M., O'Neill, B. C., Webster, M., & Agrawala, S. (2007). The limits of consensus. *Science*, 317(5844), 1505-1506.

PAI Staff. (2019). Partnership on AI Calls for Visa Accessibility Globally to Accelerate Responsible AI Development. Available at: https://www.partnershiponai.org/the-partnership-on-ai-calls-for-visa-accessibility-globally-to-accelerate-responsible-ai-development/ Accessed 21 December 2019

Pence, M. (2018) Remarks by Vice President Pence on the Administration's Policy Toward China. United States White House. Available at:

https://www.whitehouse.gov/briefings-statements/remarks-vice-president-pence-administrations-policy-toward-china/. Accessed 31 December 2019

Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Mishra, S. & Niebles, J. C. (2019). *The AI Index 2019 Annual Report*, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA.

Rawls, John.(1993) *Political Liberalism*. Columbia University Press, 1993, pp. 134–49.

Russell, B. (1945). *A History of Western Philosophy*. Allen & Unwin.

Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Penguin.

Shane, S., & Wakabayashi, D. (2018). 'The Business of War': Google employees protest work for the Pentagon. *The New York Times*, *4*.

Shead, S. (2017). Eric Schmidt on AI: 'Trust me, these Chinese people are good'. *Business Insider*. Available at: https://www.businessinsider.my/eric-schmidt-on-artificial-intelligence-china-2017-11/ Accessed 30 December 2017

Song, B. (2019). The West May Be Wrong About China's Social Credit System. *New Perspectives Quarterly*, 36(1), 33-35.

Søraker, J. H. (2006). The role of pragmatic arguments in computer ethics. Ethics and Information Technology, 8(3), 121-130.

Simonite, T. (2017). AI could revolutionise war as much as nukes. *Wired*. Available at: https://www.wired.com/story/ai-could-revolutionize-war-as-much-as-nukes/ Accessed 20 December 2019

Sithigh, D. M., & Siems, M. (2019). The Chinese social credit system: A model for other countries?. EUI Department of Law Research Paper, (2019/01).

Stewart, P. (2017). U.S. weighs restricting Chinese investment in artificial intelligence. *Reuters*. Available at: https://www.reuters.com/article/us-usa-china-artificialintelligence/u-s-weighs-restricting-chinese-investment-in-artificial-intelligence-idUSKBN1942OX Accesse 20 Decmber 2019

Sunstein, C. R. (1995). Incompletely theorized agreements. *Harvard Law Review*, 108(7), 1733-1772.

Szeghalmi, V. (2015). The Definition of the Right to Privacy in the United States of America and Europe. *Hungarian Yearbook of International Law and European Law*, 397.

Taylor, C. (1996). Conditions of an unforced consensus on human rights. Available at: http://people.brandeis.edu/~teuber/Taylor,%20Conditions%20of%20an%20Unforced%20Consensus.pdf

Tencent Research Institute, China Academy of Information and Communications Technology, Tencent AI Lab, and Tencent Open Platform. (2017). *Artificial Intelligence: A National Strategic Initiative for Artificial Intelligence* (人工知能：国家人工知能戦略的アクション). China Renmin University Press.

The Economist. (2018). How the West got China wrong. Available at: https://www-economist-com.ezp.lib.cam.ac.uk/leaders/2018/03/01/how-the-west-got-china-wrong Accessed 13 December 2019

Triolo, P., Kania, E., & Webster, G. (2018). Translation: Chinese government outlines AI ambitions through 2020. *New America, DigiChina*, 26.

US Department of Defense (2020). Release: DOD Adopts Ethical Principles for Artificial Intelligence. Available at https://www.defense.gov/Newsroom/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/ Accessed 11 March 2020

UN General Assembly (2015). Report of the Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the context of International Security. Seventieth Session, Item, 93.

Webster, G., Creemers, R., Triolo, P., & Kania, E. (2017). Full Translation: China's 'New Generation Artificial Intelligence Development Plan' . *New Ameria DigiChina*. Available at: https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ Accessed 26 December 2019

Whittlestone, J., Nyrup, R., Alexandrova, A., Dihal, K., & Cave, S. (2019) Ethical and Societal Implications of Algorithms, Data, and Artificial Intelligence: a roadmap for research. London: Nuffield Foundation.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., West, S.M., Richardson, R., Schultz, J. & Schwartz, O. (2018). AI Now Report 2018. AI Now Institute at New York University.

Yao-Huai, L. (2005). Privacy and data privacy issues in contemporary China. *Ethics and Information Technology*, 7(1), 7-15.

Ying, F. (2019). Understanding the AI challenge to humanity. China US Focus. Available at: https://www.chinausfocus.com/foreign-policy/understanding-the-ai-challenge-to-humanity. Accessed 29 December 2019.

Yunping, W. (2002). Autonomy and the Confucian Moral Person. *Journal of Chinese Philosophy 29:2, 251-268*

Zeng, Y., Lu, E., & Huangfu, C. (2018). Linking Artificial Intelligence Principles. arXiv preprint arXiv:1812.04814.

Zhang, S. (2017). China’s Artificial-Intelligence Boom. *The Atlantic*, 20170216, 20170924.

Zhang, B., & Dafoe, A. (2019). Artificial intelligence: American attitudes and trends. Available at SSRN 3312874.