【論文紹介】日本の伝統にしたがって、ほぼ全訳ですが。😅

#AI, #Model, # OpenRAIL, #responsibleAI

Carlos Munoz Ferrandis さんの「Towards open and responsible AI licensing frameworks」（Published August 31, 2022） https://huggingface.co/blog/open_rail の紹介

オープンで責任のある AI ライセンス許諾の枠組みをめざして

Open & Responsible AI ライセンス（「OpenRAIL」）は、AI 固有のライセンスで、AI 関連制作物（AI artifact）の責任のある使用を義務付ける一方で、AI 制作物へのオープンなアクセス、使用および配布を可能にするものです。OpenRAIL は、現在のオープンソフトウェアライセンスがコーディングで、また Creative Commons ライセンスが一般的なコンテンツで果たしているような役割を、機械学習（ML）の分野においてオープンで責任のある役割を果たしてくれる可能性があります：つまり、汎用的なコミュニティライセンス許諾ツールとなる可能性があります。

機械学習およびその他の AI 関連の分野での近年の進歩は、情報通信技術（ICT）分野におけるオープンソース文化の普及のおかげも一部あって、促進されてきました。そして、これは、機械学習の研究および発展というダイナミックスの中に浸透してきています。この分野でのイノベーションにとっての中核的な価値として、オープンであることの利点に争いはありませんが、機械学習モデルの発達と使用に関する倫理的および社会経済的な懸念に関連する（それほどではありませんが）最近の事態は、「オープンであることは十分でない」との明瞭なメッセージを広めています。もちろん、閉鎖的なシステムが解答ではありません。この問題は、企業内の AI 開発という透明性の乏しいプロセスについては、引き続き懸念が抱かれています。

オープンソースライセンスは万能ではない

機械学習モデルへのアクセス、開発そして使用は、オープンソースライセンスという仕組みから大きな影響を受けています。例えば、機械学習（ML）開発者達は、その開発した「重み」[1]を正式なオープンソースライセンス、または Creative Commons ライセンスのようなその他のオープンソフトウェアライセンスもしくはコンテンツライセンスを付けて提供するとき、「モデル [2]をオープンソース化する（open sourcing a model）と、よく呼んでいます。この

1 「重み」については、斉藤康毅著『ゼロから作る Deep Learning』26 ページ等、参照。
2 「モデル」という用語については、例えば、BigScience Open RAIL-M License では、“Model” means any accompanying machine-learning based assemblies (including checkpoints), consisting of learnt weights, parameters (including optimizer states), corresponding to the model architecture as embodied in the Complementary Material, that have been trained or tuned, in whole or in part on the Data, using the Complementary Material.と定義されている。

取扱は疑問を引き起きします。「なぜ、そんなことをするのか？機械学習（ML）制作物とソースコードは似ているのか？ソースコード向けに作られた私人間のガバナンスメカニズム（つまり、オープンソースライセンス）が、機械学習（ML）モデルの開発および使用にもまた適用されるべきであると技術的に十分な展望を共有しているのでしょうか？

もっとも最近のモデル開発者達は、そう考えているように見えます。オープンなかたちでリリースされたモデルの過半数には、オープンソースライセンス（例：Apache 2.0）が付けられているからです。例えば、Hugging Face Model Hub、および Munoz Ferrandis & Duque Lizarralde (2022)、参照。

しかしながら、経験的な事実は、機械学習（ML）制作物のリリースに当たって、オープンソースおよび／またはフリーソフトウェア慣行に厳格に従うアプローチ、そしてフリーソフトウェア運動の Freedom 0 原則 [3]への教条的な信仰が、機械学習（ML）モデルの使用に当たって社会倫理的な混乱をもたらしていることを、また告げています（Widder et al. (2022)、参照)。簡単に言うと、オープンソースライセンスは、ソフトウェア／ソースコードとは違った制作物としてのモデルの技フリーソフトウェア力を考慮しておらず、機械学習（ML）もであるのより責任のある使用を可能にするためには、したがって不適切なのです（例：オープンソースの定義の基準第 6)。また、Widder 等、(2022)、Moran(2021)、Contractor 等 (2020)、参照)。

個別な、その場しのぎの対応であれば、機械学習（ML）モデルのドキュメンテーション、透明性そして倫理的な利用に向けて既に図られており、日々改善されています（例：モデルカード､評価ベンチマーク)。では､なぜ､オープンソースライセンスのやり方は､機械学習(ML)モデルの個々の機能および問題点に対して、適用されるべきはないのでしょうか？

同じ懸念が、産業界および政府の機械学習（ML）ライセンス実務で生じています。Bowe& Martin (2022)の言葉を借りれば、「Anduril Industries 社のゼネラル・カウンセルである Babak Siavoshy 氏は、コンピュータビジョンによる物体検出のために産業界で開発された AI アルゴリズムについて、どのような種類のライセンス規定を適用すべきなのか、そして軍事的な照準合わせや脅威評価に､そのような技術を適用すべきかという問題を提起しました。商用ソフトウェアライセンスの規定も、あるいは標準的な DFARS データ権条項も、この問題に適切な回答を与えてくれません。どちらも、開発者の権利を保護し、また政府が責任をもって実装するためのこの AI というシステムの洞察を得るのを可能にしていないからです。

本当に機械学習（ML）モデルとソフトウェア／ソースコードが異なる制作物であるのなら、機械学習（ML モデル）がオープンソースライセンスでリリースされるのでしょうか？答えは簡単です。オープンソースライセンスは、ソフトウェアコミュニティでコードを共有する

[3] The freedom to run the program as you wish, for any purpose (**freedom 0**).

ために、ソフトウェア関連市場において事実上の標準になっているからです。協力してソフトウェアを開発するという、この「オープンソース」のアプローチが、AI 開発およびライセンス許諾実務に浸透しており、また多大な便益をもたらしているからです。オープンソースというイニシアチブと Open & Responsible AI ライセンス（「OpenRAIL」）は相互補完的なイニシアチブに十分なれる可能性があります。

それでは、なぜ、私たちは、オープンソースムーブメントに鼓舞され、また機械学習（ML）分野でのエビデンスに基づくアプローチにより導かれた一連のライセンス許諾の仕組みをデザインしないのでしょうか？実は、オープンで、責任のある機械学習（ML）の開発、使用そしてアクセスに向けた手段としての、一連のライセンス許諾の新たな仕組みが現れているのです。それが、Open & Responsible AI ライセンス（OpenRAIL）です。

ライセンス許諾のパラダイムチェインジ：OpenRAIL

RAIL イニシアチブで採用され、そして Hugging Face によって支援された OpenRAIL のアプローチは、BigScience、Open Source、そして Creative Commons のようなイニシアチブから情報の提供と刺激を受けています。OpenRAIL ライセンスの 2 つの重要な特徴は、次のとおりです。

オープンであること：これらのライセンスでは、ロイヤルティ無料のアクセスが保障され、使用許諾製品の柔軟な下流での使用および再配布、そしてそれら製品からの二次的著作物の配布の自由が保障されています。

責任のあること：OpenRAIL ライセンスは、特定された緊要なシナリオの中での使用許諾 AI 制作物の使用に関する一連の具体的な規制事項が埋め込まれています。機械学習（ML）開発に対するエビデンスに基づくアプローチから、ユーザーベースの規制事項が導かれており、また機械学習への幅広いアクセスと使用を促進することと、人工知能（AI）制作物のオープンに使用許諾の有害な使用から生ずる潜在的な社会的費用との間に一線を画することを強いる、使用法の規制が生まれます。したがって、機械学習（ML）モデルにオープンなアクセスを認めることから生ずる便益を享受する一方で、ユーザーは、指定された規制に該当するシナリオではそのモデルを使用できなくなります。

使用方法に基づく規制事項条項を、オープン AI ライセンスに統合すれば、そのモデルが乱用されたことが分かった場合に、リリースされた AI 制作物の責任ある使用を求めて立ち上がり、AI 制作物の使用をよりよくコントロールできる能力、そしてライセンサーの機械学習（ML）モデルに関する権利を強制実現できる能力が確保されます。仮にオープン AI ライセンスに行動／使用方法の規制事項がなかった場合には、ライセンサーは、その開発した AI 制作物をオープンにリリースしたら、ライセンサーが、責任のある使用方法に関する規制という法的ツールに、そもそも思い至ることができるでしょうか？OpenRAIL と RAIL は、倫理的に情報の提供される行動の規制を可能にするための第一歩です。

また、そもそも規制の強制実現について考える前に、使用法に基づく規制条項は、潜在的なユーザーが、モデルを乱用しないよう抑止する仕組みとして機能するでしょう（つまり、抑止効果です)。しかしながら、使用方法に基づく規制が存在するというだけでは、リリースされた AI 制作物の潜在的な乱用が決して起きないという保証としては不十分でしょう。これが、OpenRAIL が、AI 制作物のその後の再配布に当たり、使用方法に基づく規制を、再配布を受ける下流側に展開することを義務付け、また AI 制作物の二次的著作物のユーザーが、AI 制作物を乱用するのを抑止するための手段として、AI 制作物の二次的著作物についても規制を設けている理由です。

コピーレフト方式の行動／使用条項の効果は、その義務を、本来のライセンサーの使用許諾制作物を、責任をもって使用することに関する希望と信頼を、その要件とともに拡散させます。さらに、行動／使用方法条項が広範囲に採用された場合には、使用許諾制作物の二次的著作物をその後に配布する配布者に、その二次的著作物をコントロールする能力を与えます。社会的な視点から言うと、OpenRAIL は、当該モデルのライセンサーの設けた制限とその有する価値観を承認しつつ、AI 制作物に関する情報を提供し、尊重しながら共有するという文化の統合に向けた、ひとつの手段なのです。

OpenRAIL は、オープンソースライセンスがソフトウェアコードについて果たしてきたのと同じ優れた役割を、機械学習で果たすことができるでしょう

OpenRAIL ライセンス契約書の 3 つの例は、最近リリースされた BigScience OpenRAIL-M、Stable Diffusion の CreativeML OpenRAIL-M、そして前記の 2 つの原型である BigSicence BLOOMRAIL v1.0 （このサイトにあるポスティングと FAQ、参照）です。最後のものは、BLOOM と名付けられた BigScience の 176B パラメータモデル（と関連するチェックポイント）へのオープンなアクセスと責任のある使用を特に念頭においています。このライセンスは、一連の許諾されるライセンス条項を、大規模言語モデル（Large Language Model (LLM)）に特有のリスクと精査された制限事項のみならず、LLM の潜在的な利用形態に基づくエビデンスの根拠のある限られた数の使用上の規制が定められている、使用方法に基づく規制条項とを調和させる手法を提案することにより、オープン性と責任のある AI が交差する領域で役割を果たしています。RAIL Initiative が採用する OpenRAIL のアプローチは、また利用可能な OPT-175 や SEER のような、モデルのその他のより規制の強い行動・使用方法条項をともなうもののリリースと併行する、この種のものとしては最初のものであった BigScience BLOOM RAIL v1.0 の発展型です。

これらのライセンス契約書は、ライセンス許諾の世界での 2 つの課題に、一部対応するという BigScience の反応を象徴するものです。2 つの課題とは、(i) 「モデル」は、「コード」とは別物であること、(ii) 「モデル」の責任のある使用です。BigScience は、具体的なシナリオに基づきライセンス契約の問題点に真に焦点を絞り、そして BigScience のコミュニティのゴールを明確化して、さらなる一歩を進めています。実際に、提案されたソリューションは、

AI の世界ではまったく新しい種類のものでした。BigScience は、モデルの責任のある使用を普及させるような方法でライセンスをデザインしました（つまり、責任のある使用を推進させたのです)。その理由は、モデルの再配布または二次的著作物には、個々の使用方法に基づく規制に従わせる一方で、その他のライセンスにあっては、別のライセンス許諾条項を提案することができたのです。

OpenRAIL は、AI システムの配置、使用そして商業化について、セクター毎の個別の政府の規制を提案するという、現在継続中の規制上のトレンドにまた合致しています。政府の AI 規制の登場に合わせて（例 : EU の AI 法、カナダの AI および Data 法案)、AI の政府の規制トレンドと倫理的な懸念により示唆を受けた、新たなオープンソースライセンスのパラダイムは、来る数年で大規模に取り入れられる可能性を秘めています。オープンソースの影響、使用そしてドキュメンテーションに必要な配慮を払うことなく、モデルをオープンソース化する対応は、新たな AI の政府規制のトレンドを目にすると、懸念の源泉となる可能性があります。今後は、OpenRAIL は、進行中の AI の政府規制のトレンドを反映する実現手段と、また AI ガバナンスというより大きなシステムの一部としてのツールとして観念されなければならなりません。また、オープンで責任のある AI の使用に向けての唯一のソリューションでもありません。

オープンソースライセンスは、AI のイノベーションにとっての礎石のひとつです。ライセンスが有する社会的および法的制度としての性格は、適格に把握されなければなりません。それらは、やっかいな法律上の技術的な仕組みと捉えてはならず、むしろ、使用許諾対象の制作物をどのように使用できるかについての共通したメッセージを共有することにより、AI コミュニティの間の利害関係者を糾合するための、コミュニケーション手段なのです。

健全で、オープンで、そして責任のある AI ライセンシング文化、AI の将来のイノベーションに賭けましょう。その成果の成否は、ライセンシング、我々みんな、そしてあなたの肩にかかっています。