The Model Openness Framework: Promoting Completeness and Openness for Reproducibility, Transparency and Usability in AI

**Matt White[1,2], Ibrahim Haddad[2], Cailean Osborne[2,3] ,
Xiao-Yang (Yanglet) Liu[1,4], Ahmed Abdelmonsef[1], Sachin Varghese[1]**

[1] LF AI & Data - Generative AI Commons, [2] Linux Foundation,
[3] University of Oxford, [4] Columbia University

matt.white@berkeley.edu, ibrahim@linuxfoundation.org,
cailean.osborne@oii.ox.ac.uk, xl2427@columbia.edu,
{ahmed.abdelmonsef,sachin.varghese}@genaicommons.org

Abstract

Generative AI (GAI) offers unprecedented possibilities but its commercialization has raised concerns about transparency, reproducibility, bias, and safety. Many "open-source" GAI models lack the necessary components for full understanding and reproduction, and some use restrictive licenses, a practice known as "openwashing." We propose the Model Openness Framework (MOF), a ranked classification system that rates machine learning models based on their completeness and openness, following principles of open science, open source, open data, and open access. The MOF requires specific components of the model development lifecycle to be included and released under appropriate open licenses. This framework aims to prevent misrepresentation of models claiming to be open, guide researchers and developers in providing all model components under permissive licenses, and help companies, academia, and hobbyists identify models that can be safely adopted without restrictions. Wide adoption of the MOF will foster a more open AI ecosystem, accelerating research, innovation, and adoption.

1 Introduction

Artificial intelligence (AI) has seen remarkable advances in recent years, driven by growth in computational capabilities, increased volumes of available training data, and improved deep learning algorithms like transformers and diffusion models. Systems using deep neural networks match or exceed human performance on tasks like image recognition, speech processing, game playing, and language translation [1]. However, with this growth in capabilities, so have grown concerns regarding the transparency, reproducibility, ethics, and safety of AI systems [2], [3], [4], [5], [6].

The benefits and risks of open models have been the subject of much debate [7], [8], [9], [10], [11]. On the one hand, the accessibility and transparency of open models concurrently deliver notable advantages over closed source models, including security and performance advantages through distributed development and auditing [12], [13], adaptability and customization for diverse domains and languages [10], [14], as well as advances in the fields of science [15], [16], [17]. On the other hand, the openness of models introduces risks, such as by enabling the generation of disinformation [18], [19] or illegal content [20], [21]. According to one study, open foundational models have five distinctive properties that present both benefits and risks: broader access, greater customizability, local adaptation and inference ability, the inability to rescind model access, and the inability to monitor or moderate model usage [10]. Striking a balance between harnessing the innovation of open models and addressing associated risks remains a critical challenge in navigating the evolving landscape of AI, in particular GAI.

Most state-of-the-art (SOTA) foundation models are black boxes, making it hard to explain their internal logic or ensure they behave fairly [22]. Models are released with technical reports and model cards that provide little to no details on the source and treatment of their model's training data and subsequent fine-tuning or the methods used for model alignment. Model evaluation results often cannot be reproduced independently due to lack of disclosure of evaluation data and methods, leaving the public to largely rely on claims reported by model producers [23]. What is more, large (language) model services (LMSs) like OpenAI's GPT-4 take already opaque models and further hide them from public view, obfuscating them behind cloud-based APIs, which provide no insight into the inner workings and systems employed behind those APIs. Additionally, they do not reveal whether or not they rely on other models or systems hidden behind their APIs, leaving consumers wondering whether or not the performance they see is that of the model or secondary systems that shroud it [24]. Additionally few disclosures are made about the guardrails in place and if prompts and outputs are altered, filtered or replaced all together [25].

Following years of discussion about the safety of releasing powerful models [3], AI researchers and developers have made strides towards openness by releasing pre-trained machine learning (ML) models to the general public. Grassroots initiatives emerged as the early leaders in adopting open approaches to the development of models, such as EleutherAI, BigScience, and BigCode [26]. In recent years countless organizations and individuals have released thousands of "open" models on online platforms like the Hugging Face Hub and Kaggle, creating a rich variety of foundation and fine-tuned models available for use.

However, while the number of publicly available models has been growing, a concerning trend has emerged. Many of these models are being promoted as "open-source" but they do not utilize an actual open-source license. Adding to the confusion are fine-tuned models being released that claim to be using an open-source license like Apache 2.0 but are based on a foundation model that employs a restrictive license. This altering of the original license is not legally permitted, as derivative models always inherit the license of their foundation model. Altering a license from a restrictive foundation model to a fine-tuned open-source model is infeasible and can have legal consequences for those altering the license and those using the model.

More broadly the term "open" has been used imprecisely to describe "systems that offer minimal transparency or reusability…alongside those that offer maximal transparency, reusability, and extensibility" [27]. A review of open models found that, "while there is a fast-growing list of projects billing themselves as 'open source', 1)many inherit undocumented data of dubious legality, 2) few share the methodologies used in various stages of training including fine-tuning, and careful scientific documentation is exceedingly rare" [28].

Many of the most popular "open" models available today are billed as open source but in fact are not, a practice often referred to as "open-washing"; that is, their producers promote the models as open source but instead employ their own restrictive licenses that aim to limit downstream usage [27], [29], [30]. Similarly, some institutions claim that their models are open but then use restrictive data licenses that limit the usage of the datasets used to train their model [29]. Some model producers go so far as to add conditions that stipulate that their model outputs cannot be used to train subsequent models or add trigger conditions that would require a model consumer to negotiate a new license when some condition is met. Projects of this nature limit innovation and create risks for organizations seeking to use them for various purposes, forcing them to navigate a minefield of restrictions that make them unsuitable for broad adoption. Furthermore, many developers do not realize that open-source licenses were designed to cover conventional software code and are not appropriate for the intricacies of ML models. Open-source licenses cover the model architecture, which is defined in software code, but not the corresponding learned model parameters. Model parameters are data, as such they are not explicitly covered by open-source software licenses, instead model parameters are more aptly governed by open-data licenses [31].

Another challenge is most models have fallen short in their completeness, only releasing model architectures and final trained parameters. To achieve full transparency and reproducibility, we argue that model producers must go beyond just releasing their model and the trained weights and biases: they need to include all artifacts of their work. These artifacts include all datasets, whether training, validation, or testing, as well as those used for benchmarking. It requires the inclusion of detailed documentation, including thorough research papers, model cards, data cards, and any usage documentation. Completeness also requires all code used to parse and process data, the code used for training and inference, and any code used in benchmark tests, along with any libraries or other code artifacts that were a part of the model development lifecycle.

Various approaches have been developed by AI researchers and practitioners to address the risks posed by AI systems, including tools for auditing model explainability, fairness, and robustness [32], [33], [34], [35], [36], [37]; frameworks for assessing the risks of (open) foundation models [7], [10], [38]; and the establishment of ethics review boards within AI research labs [39], among other approaches and frameworks.

Because prior methods do not evaluate both completeness and openness of models, we seek to reinforce existing methodologies with a complementary approach — the Model Openness Framework (MOF), a system for objectively evaluating and classifying the completeness and openness of ML models based on which components of the model development lifecycle are publicly released under truly open licenses.

We also seek to differentiate between the terms “open” and “complete”, to make it clear to model *consumers* exactly what model *producers* are providing and under what conditions when they say their model is open. Openness is not just about what is included, but importantly under what licenses each component is released. We believe opening the "black box" of AI will be crucial for its continued advancement and responsible use [40].

For the sake of simplicity in nomenclature, this paper refers to any entity that develops and trains a first-generation model as a ‘model producer’ or simply a ‘producer’ This encompasses AI researchers, developers, AI hobbyists or anyone who trains a model in some form or fashion, including fine-tuning and alignment, as long as they are the originator of the foundation model. Similarly, any entity that adopts, consumes, alters or uses a model and corresponding artifacts for any purpose including modifying weights through fine-tuning is referred to as a ‘model consumer’ or simply a ‘consumer’. This includes end users, researchers, developers or anyone that uses an ML model and is not its producer. We also use the term machine learning or machine learning model to broadly describe any model, whether classical machine learning or deep learning, both generative and discriminative.

2 Understanding Completeness and Open Concepts

Before presenting the details of the MOF, it is useful to review the concepts of completeness and openness in science and technology. These core tenets form the basis of the ideals and practices of open science, open source, open data, and open access which provide the foundations for enabling transparency, reproducibility, and collaboration in research and are a part of a larger umbrella of the open knowledge movement that believes all knowledge should be shared freely [41]. This section provides an overview of each of these domains and how they connect to the goals of the framework. Understanding the motivations behind openness clarifies why it is vital to extend these concepts to AI research. Embracing tenets of openness ensures the democratization of AI which is essential to advance AI research and innovation while pursuing the pillars of responsible AI such as transparency, accessibility, and inclusivity [42].

2.1 Completeness

Completeness is the principle of releasing all key artifacts produced during the full lifecycle of conducting research or engineering a technical product to enable comprehensive transparency, inspection, evaluation, and reproducibility. In the context of ML, completeness entails releasing all the key components associated with developing an ML model rather than just selected artifacts. It entails sharing the full pipeline that produced a model in a usable form. Completeness is a core tenet of open science.

Sharing only surface elements of ML models grants only narrow visibility. Comprehensive releases empower unfettered scrutiny into model genetics: curation and treatment of training data, feature engineering, neural architectures, weight evolution, training configurations, model performance across diverse benchmarks, replication of model producer claims, and other byproducts of the model development lifecycle.

The MOF encourages model producers to exhibit full completeness, providing all artifacts involved in the model development lifecycle when distributing models. It defines an ascending hierarchy of criteria for releasing key artifacts with the highest bar aligned with open science paradigms. Completeness combined with openness (open licensing) accelerates collective advancement of trustworthy and innovative AI.

2.2 Openness

Openness is the discipline of freely sharing the methodology, progress, and products of performing research and development with the public without restrictions on access, inspection, modification, or distribution. Instead of limiting transparency through proprietary terms, openness means releasing materials using permissive open licenses tailored to the type of content. This upholds scientific ideals around reproducibility, accountability, and cumulative innovation while empowering community members to meaningfully review, discuss, reuse, and extend upon prior work. Carefully selecting appropriate open licenses facilitates attribution, protects downstream consumers, maintains community ethical norms and accelerates adoption and impact.

Within the context of ML, openness is the practice of freely sharing the artifacts and components associated with developing ML models publicly without restrictions on access, inspection, modification, or distribution. It means releasing key materials like data, code, models, and documentation under permissive open licenses instead of proprietary terms or conditional access.

Sharing machine learning models openly grants the community the freedoms to transparently review capabilities and limitations, identify issues, reuse or extend functionality, and participate in collective advancement. This is embodied through open licenses applied judiciously to raw datasets, preprocessing scripts, model architectures, trained parameters, benchmark tests, supportive code, model and data cards, research papers, and other important artifacts.

Appropriate licensing facilitates attribution, safeguards model consumers, and maintains community norms while removing barriers to adoption. It expedites the deployment of beneficial applications in the real world, preventing the concentration of power within centralized providers. The Model Openness Framework helps codify and quantify openness across model development pipelines with informative guidelines, a simple classification system and a method for assigning badges to qualified models.

2.3 Completeness vs. Openness

We use the term "completeness" borrowed from open science to disambiguate from the multiple uses of the word "openness" which has unfortunately become a vague and confusing term. Openness is often used to describe not only the licensing used for artifacts but also the availability of artifacts and even the thoroughness of those artifacts. The multiple uses of the term "open" continues to be used in a way that is misleading or doesn't reveal the specifics of its usage [43].

Packing the term "openness" with multiple definitions, uses or dimensions doesn't clearly articulate what aspect of the model is open. For instance a model producer may claim that their model is "open" but model consumers do not know if it is open because it employs open licenses, because it is made publicly available, because it provides additional components like datasets or because the components released are thorough or usable.

For this reason, we use the term "completeness" to measure the availability of components that are released with models (with the desired goal being "full completeness") and the term "openness" to describe the usage of permissive licenses for components.

2.4 Open AI

Open AI (with a space) refers to the concept of transparency and accessibility in AI research and development. This entails freely sharing key artifacts like data, code, models, and publications under both open and restrictive research licenses. This movement also includes the emerging trend of distributed development of AI models [26], exemplified by grassroots initiatives such as BigScience and BigCode [44]. This shift towards the public availability of ML models is significant, one that has caught the attention of companies that had been working on closed-source AI solutions [45].

The goals of open AI are to accelerate progress in the field through collaboration, establish trust by allowing inspection of systems, enable diverse perspectives to participate, and align AI advancement with social benefits [46]. However, it does not require that models and artifacts be released under open licenses. Due to the nascent nature of the open AI movement, the field is continually evolving and new standards are being developed to address the shortcomings including the draft of the Open Source AI Definition **[citation/reference needed]** as well as work by U.S. government agencies including NIST and NTIA. **[citation/reference needed]**

2.5 Open Science

Open science refers to the practice of making all stages of the scientific process transparent and accessible to others. This includes publishing research papers, data, source code, notebooks, and any information or tools needed to replicate research. The goals of open science are to enable reproducibility, collaboration, and facilitate building on previous knowledge to advance scientific research.

Open science has a bias towards open licenses for software, data and documentation and enables unrestricted access to research material and provides transparency into the research process, it is seen as the best methodology for putting open source, open data and open access to practice.

Open science is critical for credible, ethical, and accelerated scientific research that can be reviewed, validated, replicated and built upon. Open science in AI is sometimes referred to as "open science AI" and is the gold standard for ensuring reproducibility and transparency and is the ideal release scenario for the MOF.

2.6 Open Source

Open source involves publishing software code under licenses that grant users independence and control over the technology by allowing inspection, modification, and redistribution of the code without restrictions. The open-source development model promotes community-driven innovation and transparency. Using open-source libraries and releasing code created for research opens opportunities for contribution and extension.

Open source has emerged as an indispensable component of AI research and development, from the almost complete absence of OSS for machine learning in 2007 [47] to its widespread usage in AI research and development today [48]. Open source is also a cornerstone of open science by fostering collaboration and knowledge sharing. Embracing transparency, reproducibility, and usability without restrictions, open source encourages code and methodology sharing among researchers and developers, facilitating validation and refinement of findings. This accessibility accelerates innovation, serving as a foundation for building upon existing work and contributing to the swift progression and democratization of scientific discovery expressed in code.

The Apache 2.0 and MIT licenses are the two most frequently used open-source licenses for software and open models, both are approved by the OSI. Alternative licenses such as the Llama 2 license, OpenRAIL and AI2 ImpACT licenses are not considered open-source licenses due to the restrictions they impose on usage [30]. The Open Source Definition and the list of approved open-source licenses is maintained by the OSI [49].

2.7 Open Data

Open data is about publicly releasing the datasets, databases, and any other data used for research for access and reuse without restrictions. This upholds scientific reproducibility, allows reanalysis, and spurs innovation as well as a deeper understanding of the data used to train models. Standard policies and formats are often employed to ensure quality and usable data sharing.

There are both open-data licenses and open-content licenses, often open-data licenses are applicable to both data and content. Open content refers to creative materials (unstructured data) while open data refers to structured data.

Common open-data licenses are Creative Commons licenses, particularly Creative Commons Zero (CC0), CC BY (attribution) and CC BY-SA (Attribution-ShareAlike.) As well as the Linux Foundation's Community Data License Agreement (CDLA-Permissive) and the Open Data Commons licenses like Public Domain Dedication and License (PDDL) and the Open Data Commons Attribution License License (ODC-By).

2.8 Open Access

Open access is the process of making research outputs like publications freely available to read without subscriptions or paywalls. This enables broad dissemination of knowledge. There are various open-access platforms like Cornell University's arXiv, which make publications, often distributed under an open license, freely available for review.

2.9 Open Collaboration and Community

Open collaboration encourages cooperative efforts across institutions, disciplines, and borders. This includes partnerships between academia, industry, and independent researchers, facilitating a more inclusive and diverse development of AI technologies. Open community moves beyond open collaboration and creates a shared and safe community with neutral governance where projects can be worked on collaboratively in an equitable environment that embraces principles of openness. The Linux Foundation, LF AI & Data and the Generative AI Commons are examples of open communities [50].

2.10 Open Knowledge

Open knowledge is an overarching philosophy and larger movement that encompasses all the preceding areas of openness. It revolves around the free and public sharing of information and insights across various domains. This entails making knowledge resources accessible to everyone, contributing to a wider pool of shared understanding. Open knowledge practices also involve ensuring that the information is ethically curated and disseminated, upholding principles of integrity and respect for intellectual property.

Wikimedia Foundation, Open Knowledge Foundation and Science Commons are leading organizations in the open knowledge community.

2.11 Source Available

Source available should not be confused with open source. Source available is a paradigm that originated from conventional software development, in which a developer provides access to the source code, but the licenses of the code are not open-source licenses. This means they include restrictions that consumers of the software must fully understand before they agree to use it. Some have referred to these projects as open access, however this is a misnomer as open access applies to access to documentation without paywalls. Most open-washed projects are actually examples of source available since they use restrictive licensing.

Overall, embracing openness across science, access, data, and source code is key for credible, ethical, and accelerated research progress. The MOF aims to bring these ideals to ML research and applications allowing for a structured and informed approach to responsible innovation with reduced barriers to entry for those who work with ML models in any capacity.

3 Understanding Open Licenses

Open licenses are legal mechanisms that allow content and artifacts to be freely accessed, used, modified, and shared under permissive terms. They are essential for operationalizing open science, open data, and open-source ideals. Over the years different licenses have emerged for addressing rights, responsibilities and permissible usage for data, publications, code, and other research outputs.

For research articles and other scholarly written works, Creative Commons (CC) licenses are widely adopted. CC licenses allow free distribution and reuse with conditions like requiring attribution and allowing commercial use and derivative works. Common choices for open licenses are CC-BY (attribute) and CC-BY-SA (share alike). Using permissive CC licenses for research papers, technical reports and documentation provides rights to reproduce, expand, and translate the works.

For software code, many open-source licenses have been developed. Prominent examples include the MIT, Apache 2.0, and the 3-Clause BSD license. These licenses allow inspection, modification, and redistribution of code while requiring preservation of copyright and license terms. Using OSI-approved open-source licenses encourages community review and contributions to code, promoting quality and shared progress.

For datasets, typical licenses are Creative Commons licenses like CC-BY and CC0 (public domain), as well as data-specific licenses like CDLA-Permissive-2.0. These provide terms for sharing data openly while addressing concerns like attribution, permissive usage, and liability. Explicit open-data licensing facilitates reusing datasets for reproducibility and new applications.

Overall, open licenses solve key problems with closed, restricted systems:

- Enabling free access without paywalls or subscriptions
- Allowing reproduction, analysis, and extension of work
- Disseminating contributions back to the community
- Progressing cumulatively by building on prior ideas
- Fostering collaboration across organizational and geographic boundaries
- Promoting transparency and accountability
- Mitigating anti-competitive behavior or rent-seeking

Widespread use of open licensing for papers, data, code, and models accelerates discovery while upholding scientific norms and equity. The MOF incorporates open licenses as a fundamental component of transparent, reproducible, accessible and ethical AI.

4 The Value of Openness

Openness in software, data, research and science delivers immense value. The combination and impact of open source, open data, open access and open science is a powerful and effective way of solving the most pressing issues around AI, including increasing access, explainability, transparency, reproducibility, and innovation with the effect of enabling inspection and mitigations of the harms of model biases. This at a critical time when legislators are enacting or moving to enact legislation to control and monitor the development and usage of AI [51].

4.1 The Power of Open Source

The open-source software movement has transformed software development over the past few decades. Early closed and proprietary systems limited access, locked in users, and stagnated innovation. Opening code via licenses like Apache and MIT enabled worldwide collaborative development, freedom of choice, and accelerated progress.

Open source powers much of the Internet's infrastructure and key ML frameworks and libraries including popular projects like Linux, Kubernetes, PyTorch, and scikit-learn. Open source offers a wide range of benefits for individuals and organizations developing software, including that bugs can be found and fixed quickly, researchers and enterprises can build on open source rather than needing to start from scratch, open standards help avoid vendor lock-in, and the open-source ecosystem drives faster innovation and community collaboration. For example, a recent survey of Fortune 500 companies found that cost savings, faster development speeds, and open standards and interoperability are the primary benefits of open source for enterprises [52].

4.2 The Value of Open Data

Many fields long had cultures of data secrecy that impaired reproducibility and knowledge building. Openly sharing data enables reanalysis, reproducibility, and new applications of data.

Government open data initiatives provide transparency and spark economic activity. Opening clinical trial data improves pharmaceutical research. Public genomic databases enabled bioinformatics breakthroughs. Open geographical, social, and search data fuel innovation. Better standards and tooling around open data publishing are still needed. But the value of open data is clear.

The newly introduced Datasets and Benchmarks track at NeurIPS underscores the paramount importance of openly releasing machine learning datasets [53]. Open datasets play a crucial role in nurturing collaboration and driving innovation. They serve as a foundational resource for researchers worldwide, facilitating collective intelligence and accelerating progress in machine learning models.

Open data puts a strong emphasis on the standardization of datasets, addressing transparency and requiring specific checks, such as comprehensive descriptions of data collection methods and assessments for intrinsic bias. The commitment to accessibility is a cornerstone of open data, with datasets expected to be readily available without personal requests or paywalls, promoting transparency and enabling thorough scrutiny and well documented data origination.

4.3 The Promise of Open Access

Before open access, research publications were mostly locked behind expensive journal subscriptions and paywalls. This limited the discoverability and use of knowledge. The open access movement has made more research freely available to all.

Open access speeds the dissemination of discoveries to scientists and the public. It aids reproducibility and meta research. There are challenges in funding and transitioning established journals. But progress has occurred through open access archives, policies, mandates, and licenses. Entry barriers to accessing research have greatly reduced and providing unrestricted access to AI research papers has helped advance the field including many of the developments and enhancements to the transformer architecture that powers the latest highly capable LLMs.

4.4 The Potential of Open Science and Open AI

AI has seen explosive advancement recently through research shared via papers and conferences. But much of the training data, model details, and code remain proprietary.

The opaque nature of many AI systems limits reproducibility, slows down research, and increases concerns around bias and safety. Transitioning to open datasets, architectures, weights, and code would accelerate AI innovation and adoption for social good. Shared standards around transparency are starting to emerge. Sustained effort to open AI and adopt an open science approach will unlock its full potential not only in academia but also in the private and public sectors.

Overall, openness has repeatedly shown immense power to advance progress, equity, and opportunity across endeavors. The MOF aims to bring that spirit to AI research and practice. The AI community should learn from and join the open science in AI movement which embraces the principles of open access, open data, and open source.

5 MOF Components

The MOF defines criteria for classifying the degree of completeness and openness across all aspects of an ML model's development process including training data, model architecture, model parameters, evaluation benchmarks, and documentation. Models are classified in a 3-tier system (table 2 in Section 7) based on the specifics of which components are released openly. The higher the class indicates a more complete distribution that promotes more transparency, enabling better reproducibility, auditing, and downstream use by a model consumer and both commercial and non-commercial organizations.

The framework has 16 components to fulfill completeness of model artifacts, which cover the code, data, and documentation that are part of a typical model development lifecycle. The distribution also includes an additional component that is included with the distribution to be compliant with the MOF, the MOF configuration file:

1. Datasets
2. Data Preprocessing Code
3. Model Architecture
4. Model Parameters
5. Model Metadata
6. Training, Validation and Testing Code
7. Inference Code
8. Evaluation Code
9. Evaluation Data
10. Evaluation Results
11. Supporting Libraries and Tools
12. Technical Report
13. Model Card

14. Data Card

15. Research Paper

16. Sample Model Outputs

17. Model Openness Framework Configuration File

5.1 Datasets

Data is the lifeblood of ML models and is the most often held back element in the release of a model. Training data is data used for any form of model training including pre-training, fine-tuning, alignment using reinforcement learning techniques or data used for other methods that otherwise modify the weights of the model. Datasets also include data used for model validation and testing as well as data that may be used with benchmark tests. The datasets component may also include tokenized datasets when present.

Data can be any form or combination of media, whether text, code, images, videos, audio, 3D objects, URIs and any other data used for training, validation and testing purposes. Datasets also include any metadata, this could anything from annotation data like labels, bounding boxes and key points to attribution, bitrates, resolution and other metadata relevant to a dataset used in the model development process.

The datasets used to develop the model ideally should be released under an open license allowing unrestricted access, modification and reuse for any purpose, preferably Creative Commons CC-BY-4.0 or CC-0 but the reality is that most pre-training data is subject to copyright and it is therefore not possible to license the data. To this end as long as datasets are provided that is sufficient, whether public domain, copyrighted data or any form of license. Having access to the training data, whether pre-training, fine-tuning, alignment or any other data enables reproducibility and validation of the training process.

Any limits on sharing due to privacy or sensitivity should be documented. It is ideal that both pre and post-processed data be supplied, however with the size of many datasets, providing links to any curated raw datasets online is sufficient when accompanied by data preprocessing code.

5.2 Data Preprocessing Code

The data preprocessing code is all code used for preprocessing, cleaning, and formatting the training, validation and testing data for a model. It may also include code used to transform fine-tuning data and that used for alignment tasks like Reinforcement Learning from Human Feedback (RLHF). Other data preprocessing code could include code for data ingestion when appropriate, feature engineering, data augmentation and tokenization. Sharing this code aids in reproducibility and can help uncover data-related biases.

The data preprocessing code must be released using an OSI-approved license that covers open-source software.

5.3 Model Architecture

The model architecture is the core of any ML project – it can include the ML algorithms, neural network layout, connectivity, activations and other architectural elements. The model architecture should be fully described in the paper and shared as open-source code. This enables implementation, analysis, extensions, adaptations and unrestricted usage of the model or models.

The model architecture is a code artifact and to be considered open, must be released under an OSI-approved open-source license that does not limit its usage and derivative works.

5.4 Model Parameters

Trained model parameters must be released under an open license allowing unrestricted usage, and allow for the study, modification and redistribution of model weights. In the case of deep learning models, checkpoints from key intermediate stages of training as well as the final optimizer state ideally should be included to enable reconstructing the full model lifecycle for complete reproducibility. However at a minimum the final model parameters and optimizer state (when applicable) must be distributed in an acceptable format whether compressed or uncompressed for usage with popular deep learning frameworks like TensorFlow, Keras, PyTorch or in the framework independent ONNX file format.

Recently many model producers have released model parameters using an open-source license like Apache 2.0 and MIT, however model parameters (weights and biases) are not compatible with open-source software licenses. Since model parameters are in fact data, producers should use an open data license like CDLA-Permissive-2.0. Although licenses designed for open-source software are permissive and indemnify the developer from any liability, open data licenses are better suited to data specific considerations such as privacy, ethics, and data rights. Most permissive software licenses do not refer to data directly and do not address the ability to modify and redistribute model parameters. This gap could result in a legal obligation to any model consumer if the model producer were to implement royalties after widespread adoption of their model. This is a legal gray area that remains untested.

Ideally the model architecture and the model parameters should be saved independently in different files for distribution as each one requires a different format-appropriate open license, and each component can be studied, modified, redistributed and used independently of the other.

5.5 Model Metadata

There are other forms of metadata that can provide additional context about the model, such as the version of the framework used to create it and custom tags or descriptions provided by the developer including model and data lineage information. There is no particular requirement or profile for this type of metadata and it can reveal anything the developer would like to include with the shipped model. This information can help with model management, especially when working with multiple versions of models or conducting experiments. Often the metadata is exported from or loaded by a metadata store.

Any model metadata should use an open-data license such as CDLA-Permissive-2.0.

5.6 Training, Validation and Testing Code

The full code for training, validating and testing the model, including model construction, training loop, hyperparameter selection, and checkpointing should be open sourced. Any fine-tuning code, reinforcement learning code or any other method that otherwise modifies that model parameters or code that implements adapters that ultimately affect the performance of the model must be included. This enables reproducible training from end-to-end. Comments explaining the approach should be included in the code, ideally following PEP 8 style guide for Python code. In addition, the inclusion of log files generated during training also provides deeper insights and are recommended to be included with a release.

The training, validation and testing code itself must be released under an OSI-approved open-source license while any logs should use a permissible open-content license like CC-BY-4.0.

5.7 Inference Code

Code for performing inference with the trained model must be shared under an open-source license. This includes any data preprocessing or postprocessing required during inference. It can include any model optimizations and dependencies like external libraries. It fundamentally includes any code required to fully replicate the benchmark results presented in the research paper for the project. Availability of this code facilitates complete replication of the performance of the model and informs the consumer how to use the model most effectively for their applications.

The inference code must be released under an OSI-approved open-source license.

5.8 Evaluation Results

Detailed quantitative metrics and qualitative results from evaluating the model must be reported in the research paper or the technical report. Tests can evaluate any factor, not limited to model efficiency, accuracy, performance, fairness and bias evaluations, toxicity, truthfulness and so forth. Producers must include benchmark test results, whether industry standard benchmarks or custom benchmark tests that were developed. If industry standard benchmark tests or test suites are used, the test suite name, test name and version number must be included with the results. If custom benchmarks were developed, whether in code or any form of media including text, images, the custom benchmarks must be included in full for validation.

The evaluation results should be summarized in the technical report and the research paper, depending on which class of the MOF the distribution is released under. The raw outputs of the model evaluation should be distributed for easy verification by model consumers and use an open license suitable for content like CC-BY-4.0.

5.9 Evaluation Code

Evaluation code, evaluation data and evaluation results are separate components in the MOF. This is due to the fact that some benchmarks are written in code and other benchmarks only use data, for instance text used to evaluate an LLM or images used to evaluate a computer vision model. Many benchmark tests are a combination of both code and data used to evaluate a model, which includes the scripts needed to load the data and run benchmark tests. Since code and data require different licenses, they are separate components. Depending on the nature of the model and the methods used to evaluate it, the distribution may include one or both of evaluation code and data.

Any code used for model evaluation and benchmarking must be included and distributed under an OSI-approved open-source license.

5.10 Evaluation Data

In the case where the model is evaluated with data, that data being any media format including text, images, videos, audio, 3D data, and so forth, then that evaluation data must be included with the distribution. In the event that the model is not evaluated with data, then the evaluation data is not required.

Where the model producer relies on standard benchmark tests that are widely disseminated, it is not necessary to include them with the distribution, but they are described in the technical report and whitepaper, along with the version of the test.

If included in the distribution, the evaluation data must use a data or content appropriate permissive license like CDLA-Permissive-2.0, CC-BY-4.0 or CC0.

5.11 Supporting Libraries and Tools

Releasing any supporting code libraries, utilities, or tools developed in the course of the research under an open-source license makes them available for wider use. This could include data loaders, visualization code, simulation environments, etc. Use of existing and custom open-source tools should also be documented.

Other tools and libraries may include any of the following:

- Software libraries and frameworks used in model development along with version details.
- Tokenizers – Code used to tokenize text and any data used to train the tokenizer (if used.)
- Hyperparameter search code - Code for automating hyperparameter tuning (if used).

- Compute infrastructure code - If specialized compute infrastructure was built to scale training, the setup code could be released.
- Monitoring code - Code for tracking experiments, metrics, artifacts etc. during model development is often useful to open source as well.
- Containerization files - Dockerfiles or other container packaging to distribute the model could be shared.
- Frontend/visualization - Any web/mobile frontends or visualizations built on top of the model outputs could be released as open source.
- Deployment orchestration - Infrastructure-as-Code templates for deploying the model to production.
- Model integration code - Wrapper code/SDKs to integrate the model into downstream applications.
- Interactive demos - Links to hosted interactive demos of the model through Jupyter, Streamlit, etc.

Presumably most libraries and tools will already have their own licenses, so only in the event that the model producer creates their own libraries or tools would they need to include them with the distribution and use an OSI-approved license for the software.

5.12 Model Card

A model card provides metrics, usage guidance, and details about a model. Model cards should cover model details, intended uses, factors, evaluation, risks, and mitigations related to the model [54]. This provides transparency into model behavior.

The model card itself must use a permissive license that covers documentation, ideally CC-BY-4.0.

5.13 Data Card

A data card provides summary statistics and other details about a dataset to better understand its composition. Following guidelines like the Data Nutrition Project, data cards should describe metrics about the features, instances, intended uses, motivation, and collection process. They help identify data biases and steer proper usage and prescribe the process for reproducibility and transparency into the entire data preparation process.

The data card must be released under a permissive license that covers documentation, ideally CC-BY-4.0.

5.14 Technical Report

The technical report is a document that is much less detailed than a research paper. The technical report provides the necessary documentation for the model consumer to understand the performance, usage and implications of using the model but not enough details to reproduce the model and replicate its results. The technical report is not necessary if a research paper is included, but can be included in the distribution. The goal is to characterize model capabilities and provide guidance for successful adoption and impact.

The technical report must be released under an appropriate open license for documentation, ideally CC-BY-4.0 or CC0. It should be made available on an open access platform free of paywalls but it must be included in the distribution for permanence.

5.15 Research Paper

The paper detailing the model methodology, results, and analysis should follow principles of open science. Releasing under a Creative Commons license, preprints, and open access maximizes accessibility and transparency of the research. We suggest following a structure like: abstract, introduction, related work, methods, results, discussion, conclusion, references. However, we are not prescribing a specific format as a part of the MOF.

The research paper must be released under an appropriate open license for documentation, ideally CC-BY-4.0 and must be shared on an open-access platform such as arXiv and included in the distribution.

5.16 Sample Model Outputs

Sample outputs generated by the model are not required for the MOF, but if they are included in the distribution, they must be shared publicly without copyright or restrictions where legally permitted so that they can be redistributed with the release. These outputs could be text samples, images, videos, software code, audio, 3D assets, metadata or any other potential output generated from the model including predictions and probabilities.

For certain sensitive domains, generated examples can be anonymized or simulated if needed. Sample model outputs assist others in performing a quick evaluation of the performance of the model and a glimpse at its capabilities. In the event where model outputs are not copyrightable, the outputs should be released without a license, and this should be noted in the LICENSE file. Sample model outputs are not required for the MOF and are rather a recommendation.

Actual model outputs that are generated once the model consumer performs inference are not considered in the MOF.

5.17 Model Openness Configuration File

The MOF configuration file is an important element of any distribution and must be included so that model consumers and platforms that host models know what model components are included in the release and what licenses each component uses.

The file itself is distributed under the Creative Commons CC-BY-4.0 license. More on the MOF configuration file appears later in this document.

6 Model Openness Framework Acceptable Licenses

Each component is categorized into a domain of either Data, Model, or both. The content type of each component is also classified as data, code, or documentation. Finally, we specify standard open licenses that should be used for releasing that component - with some flexibility to allow equivalent licenses.

This comprehensive scope encourages opening the entire pipeline that produces, evaluates, and applies a model, providing multiple views into its inner workings. Table 1 summarizes this structure.

Component	**Domain**	**Content Type**	**Accepted Open License**
Datasets	Data	Data	Preferred: CDLA-Permissive-2.0, CC-BY-4.0 Acceptable: Any including unlicensed

Data Preprocessing Code	Data	Code	Acceptable: OSI-approved
Model Architecture	Model	Code	Acceptable: : OSI-approved
Model Parameters	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: OSI-Approved, Permissive Open Data Licenses
Model Metadata	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: CC-BY-4.0, Permissive Open Data Licenses
Training Code	Model	Code	Acceptable: OSI-approved
Inference Code	Model	Code	Acceptable: OSI-approved
Evaluation Code	Model	Code	Acceptable: OSI-approved
Evaluation Data	Model	Data	Preferred: CDLA-Permissive-2.0 Acceptable: CC-BY-4.0, Permissive Open Data Licenses
Evaluation Results	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses

Supporting libraries and Tools	Model	Code	Acceptable: OSI-approved
Model Card	Model	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Data Card	Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Technical Report	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Research Paper	Model & Data	Documentation	Preferred: CC-BY-4.0 Acceptable: Permissive Open Content Licenses
Sample Model Outputs	Model	Data or Code	Unlicensed

Table 1: Model Openness Framework Components and Licenses

7 Model Openness Framework Classes

7.1 MOF Structure

The MOF divides ML components into 3 separate classes, each building upon the previous. This approach is much more meaningful than a calculated index, as it directs model producers to provide prescribed essential components released under open licenses for each tier of the framework. With each increase in class of the MOF, the producer moves closer to a more complete distribution that best mirrors open science in AI.

Table 2 outlines the three classes of the MOF, each class building upon the previous. To qualify for a particular class, the producer must provide every required component in that class and each component must be released using an appropriate open license from table 1 to qualify the entire project at a particular class level. Class III being the least complete, class II being more complete and Class I being the most complete.

MOF Class	**Components Included**
Class I – Open Science	● Research Paper ● Datasets ● Data Preprocessing Code ● Model Parameters *(Intermediate Checkpoints and Optimizer States, log files)* ● Model Metadata ● And all Class II Components
Class II – Open Tooling	● Training Code ● Inference Code ● Evaluation Code ● Evaluation Data ● Supporting Libraries & Tools ● And all Class III Components
Class III – Open Model	● Model Architecture ● Model Parameters (*Final Checkpoints and Optimizer State*) ● Technical Report ● Evaluation Results ● Model Card ● Data Card ● Sample Model Outputs

Table 2: Model Openness Framework Classes and Components

7.2 MOF Class Descriptions

The 3 classes of the MOF represent ascending levels of model completeness and openness. We describe the distinguishing aspects of each tier beginning with the lowest class.

Class III - Open Model

In the MOF, Class III is the entry point and contains the minimum required components that must be released using open licenses. If not all of these components are included in a release and all components do not use an open license then the entire release cannot be considered open under the MOF. The Open Model class covers the following:

- Core model architecture and the final set of parameters
- Light documentation conveying capabilities and characterization of the model and data.

This class contains all of the components required for any model consumer to study, modify, redistribute and build upon a model, including using the model for commercial, educational and any other viable purpose without restrictions. The inclusion of the model architecture and the final weights and biases, along with documentation in the form of a technical report, evaluation results (which may be included in the technical report) along with the model and data cards provides the necessary information to work with the model and understand its capabilities, constraints and the nature of the data used to train it but not the specifics of the data.

However this class does lack completeness and robustness for full reproducibility, as well as the level of transparency needed to confirm all claims made by the model producer. It also lacks sufficient components to evaluate the model at a deep level including analyzing the training data.

Class II - Open Tooling

Building upon Class III, level II provides model consumers the complete codebase including libraries and tools needed for training, assessing and testing models themselves. Added elements include:

- Full training and inference code
- Benchmark tests to validate and quantify performance
- Libraries and tools to ease integration and to complete the codebase

This tier is an intermediate step between an open model and open science, providing a module consumer with all that is needed to test model producer assertions like benchmark results. It also allows a model consumer to perform debugging, and allows for enhancements to model functionality.

Although it does provide insights into the training process it does not include the actual datasets. It is also lighter on documentation which hampers a deeper understanding of the model's intricacies.

Class I - Open Science

The top tier aligns with ideals of open science - sharing all artifacts needed for end-to-end transparency and collaboration. This includes:

- A detailed research paper conveying the genesis of the model and its evolution
- Raw training datasets used in the training of the model
- Checkpoint weights showcasing full model evolution
- Log files providing yet more low-level insights

Fulfilling Class I empowers the community to inspect models through the model lifecycle along multiple fronts. It represents the gold standard for completeness and openness, spurring cumulative innovation rooted in scientific principles.

7.3 Hybrid Releases

Dating back to the early days of the open-source movement, openness has always been a binary decision, either software is open-source or it is not, there is no in-between or gradient approach. Either a developer released their software under an OSI-approved license or they did not. When some component of the software was not released under an open-source license, and that component was essential to the usability of the software, then the entire release was no longer considered open source.

The MOF follows this same principle. When any component is not released using an open license as described in Table 2, then that component is not deemed to be open and does not qualify for a class in the MOF. Where the removal of a component would then move the project into a lesser class this is acceptable as long as all components that are released in the prescribed class are released with open licenses.

To qualify as a Class III project that is minimally open, the model, its parameters and a technical report that describes the work along with evaluation results and model and data cards must be released with open licenses. If this is not the case, the project cannot be considered open. This includes all projects that use modified open licenses and implement any form of restrictions or acceptable uses.

It should be noted that the MOF classifies models and their components on completeness when they are open. The reader should not confuse the classification system as being a gradient measure of openness, but rather a measurement of the completeness of a release towards open science principles.

8 Implementing the Framework

8.1 MOF Process Overview

Unlike other frameworks that attempt to dictate how model producers should build and train their models or create a release path on how models should be released, we take a more objective approach and evaluate models based on their completeness and openness. This approach does not constrain model producers into a single methodology but rather lays out a pliable process that acts as a guideline to help model producers create the most complete and open models. At the completion of the process the MOF provides model producers with a badge for their MOF class that clearly demonstrates to the public their commitment to both completeness and openness.

The MOF process generally follows these steps:

1. **Inventory artifacts**
    - Comprehensively list all artifacts involved in creating the model - data, code, documentation etc.
    - Capture details like component names, component locations, versions and licenses.
2. **Map to MOF components**
    - Align inventory items to the 16 components defined in Section 5.
    - Multiple inventory elements may map to a single standard component.
3. **Verify licenses**
    - For each MOF component present, check if it uses an acceptable open license from Table 1.
    - If licenses are incompatible, the model cannot be classified.
4. **Determine completeness**
    - Check inventory against the component list for the 3 classes in Table 2.
    - Classify model at the highest tier where all required components in the class employ open licenses.
    - Model meets Class III at a minimum when using open licenses.
5. **Generate MOF.JSON**
    - Create the MOF.JSON file in the format illustrated in Figure 1, either using the Model Openness Tool (MOT) or manual means.

- Include all artifacts, licenses, locations and other required data to meet the MOF requirements.

6. **Self-assert classification**
    - With inventory, mapping, and MOF.JSON file finalized, the model producer asserts the appropriate class using the MOT or through self-assessment.
    - The model producer must stand behind their completeness and openness claims.

7. **Badging and validation**
    - The model producer uses the MOT for badging classified models.
    - MOT provides the MOF.JSON file and badge code for inclusion with project files.
    - Community helps ensure accurate labeling by filing disputes.

This process determines a model's location on the spectrum - guiding model producers in improving openness and consumers in evaluating fitness of models for their usage.

8.2 Preparing the Distribution

All projects must include a LICENSE file that describes the licenses used for the project. Conventionally a LICENSE file would include a single license, however it is recommended that the LICENSE file include all licenses that apply to the project. For instance if software is covered under Apache 2.0 and all documentation and data use CC-BY-4.0, then the text of both licenses should be included in the LICENSE file in their entirety including the license heading in order to distinguish what text belongs to which license. Alternatively, a distribution can contain different LICENSE files that are bound to the different components included in the distribution. Ideally the LICENSE files for each component should be located in the base directory of the component that they cover. The MOF.JSON file records the path to the appropriate LICENSE file for each component included in the distribution and facilitates both the per component LICENSE method and the single LICENSE file method.

In addition to the LICENSE file, the distribution must include an MOF.JSON file to provide details about the MOF version, details associated with the release, describe the included components and the licenses that each component uses. Below is an example of a MOF.JSON file for a project that is a Class I and includes all components with open licenses. This file can be generated with the MOT maintained by the Generative AI Commons at https://isitopen.ai or can be created manually or by some other automated fashion.

It is important to note that when a component is not released with the distribution, then it should not appear in the MOF.JSON file. When a component is released but does not use an open license or it uses a custom license, it should not be included in the MOF.JSON file either. The MOF.JSON file only references included components that are released using an open license.


```
{
  "framework": {
        "name": "Model Openness Framework",
        "version": "1.0",
        "date": "2024-12-15"
  },
  "release": {
        "name": "GPT-Z-instruct",
        "version": "7B",
        "date": "2024-01-30",
        "type": "language model",
        "architecture": "transformer",
        "treatment": "instruct fine tuned",
        "origin": "GPT-Z",
        "producer": "Generative AI Commons",
        "contact": "contact@isitopen.ai",
        "mof_class": 1
  },
  "components": {
        "Datasets": {
        "description": "Training, validation and testing datasets used for the model",
```

```
"location": "/path/to/datasets/",
"license_name": "CDLA-Permissive-2.0",
"license_path": "/path/to/datasets/LICENSE"
},
"Data Preprocessing Code": {
  "description": "Code for data cleansing, normalization, and augmentation",
  "location": "/path/to/preprocessing_code/",
  "license_name": "Apache-2.0",
  "license_path": "/path/to/preprocessing_code/LICENSE"
},
"Model Architecture": {
  "description": "Well commented code for the model's architecture",
  "location": "/path/to/model_architecture_code/",
  "license": "Apache-2.0",
  "license_path": "/path/to/model_architecture_code/LICENSE"
},
"Model Parameters": {
  "description": "Trained model parameters, weights and biases",
  "location": "/path/to/model_parameters/",
  "license": "CDLA-Permissive-2.0",
  "license_path": "/path/to/model_parameters/LICENSE"
},
"Model Metadata": {
  "description": "Any model metadata including training configuration and optimizer state",
  "location": "/path/to/model_metadata/",
  "license": "CDLA-Permissive-2.0",
  "license_path": "/path/to/model_metadata/LICENSE"
},
```

```
"Training Code": {
    "description": "Code used for training the model",
    "location": "/path/to/training_code/",
    "license": "Apache-2.0",
    "license_path": "/path/to/training_code/LICENSE"
},
"Inference Code": {
    "description": "Code used for running the model to make predictions",
    "location": "/path/to/inference_code/",
    "license": "Apache-2.0",
    "license_path": "/path/to/inference_code/LICENSE"
},
"Evaluation Code": {
    "description": "Code used for evaluating the model",
    "location": "/path/to/evaluation_code/",
    "license": "Apache-2.0",
    "license_path": "/path/to/evaluation_code/LICENSE"
},
"Evaluation Data": {
    "description": "Data used for evaluating the model",
    "location": "/path/to/evaluation_data/",
    "license": "CDLA-Permissive-2.0",
    "license_path": "/path/to/evaluation_data/LICENSE"
},
"Evaluation Results": {
    "description": "The results from evaluating the model",
    "location": "/path/to/evaluation_results/",
    "license": "CC0-1.0",
```

```
    "license_path": "/path/to/evaluation_results/LICENSE"
  },
  "Supporting libraries and Tools": {
    "description": "Libraries and tools used in the model's development",
    "location": "/path/to/libraries_and_tools/",
    "license": "Apache-2.0, MIT",
    "license_path": "/path/to/libraries_and_tools/LICENSE"
  },
  "Model Card": {
    "description": "Model details including performance metrics, intended use, and limitations",
    "location": "/path/to/model_card/",
    "license": "CC-BY-4.0",
    "license_path": "/path/to/model_card/LICENSE"
  },
  "Data Card": {
   "description": "Documentation for datasets including source, characteristics, and preprocessing details",
    "location": "/path/to/data_card/",
    "license": "CC-BY-4.0",
    "license_path": "/path/to/data_card/LICENSE"
  },
  "Technical Report": {
    "description": "Technical report detailing capabilities and usage instructions for the model",
    "location": "/path/to/technical_report/",
    "license": "CC-BY-4.0",
    "license_path": "/path/to/technical_report/LICENSE"
  },
  "Research Paper": {
    "description": "Research paper detailing the development and capabilities of the model",
```

```
            "location": "/path/to/research_paper/",
            "license": "CC-BY-4.0",
            "license_path": "/path/to/research_paper/LICENSE"
        },
        "Sample Model Outputs": {
            "description": "Examples of outputs generated by the model",
            "location": "/path/to/sample_outputs",
            "license": "Public-Domain",
            "license_path": ""
        }
  }
}
```


Figure 1: Sample MOF.JSON file

MOF.JSON Structure:

The MOF JSON file is structured as a single MOF object defined at the root of the JSON file, under the root there are 3 required nested objects with their own set of variables, they are:

- **Framework:** This object contains the details related to the framework itself, including the following required variables:
  - **name**: name of the framework. The variable type is string.
  - **version**: version number of the framework. The variable type is string.
  - **date**: date the framework version was published. The variable type is string in the format "YYYY-MM-DD".

- **Release:** This object contains the details of the model being released. There are a number of variables:
  - **name:** the name of the release. The variable type is string.
  - **version:** the version for the release, which can be the parameter count or some other identifier that distinguishes the model from both

previous versions and versions of the same model with different parameter counts. The variable type is string.

  - **date**: the date of the release. The variable type is string in format "YYYY-MM-DD".
  - **type**: the nature of the model, i.e., language model, image generation, audio generation, image classification, statistical ML, or any number of other types of models. The variable type is string.
  - **architecture**: the model architecture employed, i.e., transformer, diffusion, GAN, NERF, VGG, Resnet, K-means, or any other type of model architecture. The variable type is string.
  - **treatment**: any type of post-training treatment, like fine-tuning, constitutional alignment, RLHF or any other treatment that otherwise modifies the parameters of the original model. If no treatment has been applied then this variable is an empty string. The variable type is string.
  - **origin**: the original model, generally this is the foundation model. If this is not a foundation model in the release, then this variable contains the name and version of the model that was modified. The variable type is string or left empty for foundation or non-derivative models.
  - **producer:** the name model producer or publisher, could be a company, organization, group or individual. The variable type is string.
  - **contact**: an email address for the model producer or publisher. The type is string, the format is an email address.
  - **mof_class**: the qualifying Model Openness Framework class of the release as generated by the Model Openness Checker. The variable type is integer.

- **Components:** This object contains a list of components that are included with the model distribution, as well as each component's details:

  - **description**: A text description of the component, using the default values is acceptable. When introducing a new component that is outside of the standard components it is important to include a description of the component.
  - **location**: the location of the component within the distribution, full path is required in UNIX format with leading slash for the root directory. The variable type is string.

- ○ **license**: the SPDX identifier of the license or licenses used for the component. In the case where multiple licenses are used for a single component, which is often the case for libraries and tools, they must be provided in a comma-separated list. The license field must use a valid SPDX license identifier, found here: https://spdx.org/licenses/. The variable type is string.
- ○ **license_path**: the location of the LICENSE file for the component within the distribution, full path is required in POSIX format with leading slash for the root directory. More than one component can point to the same LICENSE file. In the event the component employs multiple licenses, the LICENSE file should contain the text for all the licenses used. Alternatively, multiple license files may be specified, each separated by a comma. However they must correspond in order to the comma separated list of license names provided in the license variable. The variable type is string.

8.3 Class Assignment

The MOF relies on self-reporting and projects are not classified by a central authority. LF AI & Data Generative AI Commons will provide a web interface, the MOT, that allows model producers to fill out a web form with the details of their project and in turn the MOT informs the user how their project lines up with the classes of openness in the MOF.

8.4 Badging System

The MOF is designed to be both informational and actionable. As such the Generative AI Commons is implementing a badging program, similar to the OpenSSF Best Practices Badge Program (https://www.bestpractices.dev). The badging system is a part of the MOT (https://isitopen.ai), and is a free service that allows model producers to perform the following:

- Perform a check the completeness and openness of their model distribution and display which MOF class their model meets
- Receive recommendations on which licenses to use for which components
- Generate an MOF.JSON file for their distribution
- Be provided with code to insert into their README.md file in their Github repo
- Track their model's ranking amongst other models on the MOF scoreboard

For model consumers, they can do the following:

- View the MOF scoreboard to see which models are the most complete and open
- Drill down into model distributions to see which ones meet their completeness and openness requirements
- Quickly see which MOF class a model has attained in the project's Github repo
- Validate that a model has attained an MOF class
- Submit a dispute if they believe that a model is being misrepresented as complete or open

It is incumbent upon the producer of an ML model and its components to accurately include the results of either the MOT or accurately identify the components and licenses included in the distribution in the MOF.JSON file and specify the class the project qualifies for. Misrepresentations will only harm the reputation of the model producer.

8.5 Disputes

The MOF relies on the honesty and transparency of the community to accurately classify their ML models and clearly state which components they include and the licenses that each component implements. Therefore, we also rely on the community to identify projects that have been misrepresented as open and notify the organization that hosts the project about their concerns.

9 Benefits of the MOF

Adopting the MOF brings many advantages:

- **Clarity** – Clearly defines what components are included and under which license each is distributed, in order to understand the acceptable forms of use and whether a project is complete and truly open or not.

- **Openness** – By classifying models and their artifacts at increasing degrees of openness, the MOF will help push model producers towards creating the most complete and open models, helping to advance open science and both academic and commercial usage.

- **Reproducibility** – Comprehensive availability of data, code, and models enables others to independently reproduce results and identify sources of errors, bias or disparities. This strengthens scientific rigor.
- **Transparency & Explainability** – Opening model architectures, weights, training code, and documentation sheds light on how models work and behave. This builds appropriate trust and aids inspecting for issues.
- **Data Provenance** – Origination and attribution can be determined when the data and its details are released. This can be helpful in tracing bias in models or identifying sources of PII leakage.
- **Accountability & Fairness** – Public data and models can be audited for unwanted biases and harms. Model producers can be notified of problems discovered by the community.
- **Continuous Improvement** – Model producers and consumers can build on open models instead of starting from scratch, accelerating innovation and progress in AI.
- **Collaboration** – Sharing open resources allows model producers and consumers across different fields and organizations to pool knowledge and capabilities.
- **Education & Learning** – Data, code, and models support teaching and learning about AI. Students and new researchers and developers can more easily enter the field.
- **Regulation** – Openness makes models more amenable to oversight and governance, unlocking public policy options.

Widespread adoption of the MOF will drive faster and more responsible advancement of AI. It establishes completeness and openness as primary criteria alongside the core tenets of responsible AI. The larger AI community should recognize and reward complete and open distribution of models. Model producers should incorporate the framework into their policies to make open science the standard for model distribution.

10 Limitations and Criticisms

10.1 Known Limitations

We acknowledge several limitations and likely criticisms:

- The MOF is designed around deep learning artifacts but does not transfer directly to every form of learning in AI. The framework is applicable to classical ML but does not entirely translate well to all aspects of reinforcement learning.
- Model producers are expected to be honest about the availability of the components released with their models and the openness of licenses for each component as well as the completeness of both in their release.
- Requires convincing model producers who may be reluctant to share their work publicly without restrictions initially.
- Openness goals must be balanced with privacy, IP, institutional policies, and commercialization pressures.
- Classifying models ignores their actual functionality, and bias, safety, and other harms remain a concern. However, openness with models and data enables external audits of quality and completeness.
- Simplicity of classification may not capture all nuances. But enhancement of the rubric may occur.
- Does not address the use of copyrighted materials included in training data. This is an area that is currently being addressed through the courts and legislation. It is sufficient for the MOF that data be open by using an open license, however, we encourage model producers to use data they are authorized to use in the training of models and respect copyrights [55].

11 Related Work

Our framework builds on preceding ideas such as the AAAI Reproducibility Checklist [56] and The NeurIPS 2019 ML Reproducibility Checklist [57]. The key novelties are comprehensively covering all parts of the model development pipeline, integrating with open licenses, and quantifying model openness, and presenting in an easy to identify hierarchy of completeness and openness that encourages producers to strive for complete transparency and usability without restrictions.

The authors are aware of a paper on a gradient approach to model releases where the author categorizes models like Bloom by BigScience as “open” alongside EleutherAI’s GPT-J. The authors of this proposal view open models as those that are released with licenses that do not impose downstream restrictions, so we see openness as a binary attribute, where completeness measures the dimension of component availability. Models that implement restrictive licenses are best described as source available models. EleutherAI’s GPT-J is an example of an open model as it was released under the Apache 2.0 license which is an OSI-approved license while Bloom was released under a restrictive, non-OSI-approved license, OpenRAIL [58], making it a source available model.

The OSI is currently working on defining the term “Open Source AI” [59] which is out of the scope of this proposal. Although open-source licensing is an absolute imperative for the components that are provided in code for the MOF, we see the MOF as being aligned with open science principles and the original vision of open AI, which requires more than open-source licenses to be considered open. Non-code elements like datasets and research papers need an appropriate license that suits its format, such as open-data or open-content licenses, which are not currently OSI approved licenses.

12 Out of Scope

The MOF is not designed to solve all issues related to AI and openness and relies heavily on the community to be transparent and honest in the reporting of the components they release and the licenses applied to each.

The MOF does not intend to address any of the following as they are best addressed through alternative methods, other industry activities or the courts:

- Bias and fairness
- AI safety
- Trustworthiness
- Performance testing
- Red-teaming
- Security and privacy
- Components related to model serving

- Model provenance

13 Conclusion

The MOF provides a clear, actionable methodology for assessing and improving the transparency of ML models. By outlining specific components that should be openly released across training data, code, model architecture, model parameters, documentation, and more, it gives model producers a roadmap to follow for reproducible and ethical AI development.

Adopting open licenses, as prescribed by the framework, enables collaboration, community and the freedom to use, modify, and distribute models and components under the terms of its license. The tiered classification system incentivizes releasing models with increasing levels of completeness. Widespread use of the framework would accelerate AI progress through collective innovation while ensuring fairness, safety, and public oversight.

Realizing this vision requires a concerted effort by all AI stakeholders - researchers, developers, institutions, companies, governments - to embrace both completeness and openness as core tenets. But the immense benefits for science, business, and society make pursuing model transparency well worth the challenge.

With carefully designed incentives, policies, and community norms, open source and open science ideals can become the norm in AI, not the exception. By working together across domains, we can shape AI advancement to be as complete, open, ethical, and empowering as possible. The MOF provides practical guidance for this journey towards trustworthy and democratized AI.

References

[1] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[2] Hutson, M. (2018). AI Researchers Allege That Machine Learning is Alchemy. Science, 360(6388), 860-861.

[3] Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., Radford, A., Krueger, G., Kim, J.W., Kreps, S. and McCain, M. (2019). Release strategies and the social impacts of language models. arXiv preprint arXiv:1908.09203.

[4] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR. http://proceedings.mlr.press/v81/buolamwini18a.html?mod=article_inline&ref=akusion-ci-shi-dai-bizinesumedeia.

[5] Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A. and Kenton, Z.. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359.

[6] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big?🦜. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623). https://dl.acm.org/doi/abs/10.1145/3442188.3445922.

[7] Solaiman, I. (2023). The Gradient of Generative AI Release: Methods and Considerations. arXiv. https://doi.org/10.48550/arXiv.2302.04844.

[8] Law, H., & Krier, S. (2023). Open-source provisions for large models in the AI Act. https://www.repository.cam.ac.uk/handle/1810/354175.

[9] Seger, E., Dreksler, N., Moulange, R., Dardaman, E., Schuett, J., Wei, K., Winter, C., Arnold, M., Ó hÉigeartaigh, S., Korinek, A., Anderljung, M., Bucknall, B., Chan, A., Stafford, E., Koessler, L., Ovadya, A., Garfinkel, B., Bluemke, E., Aird. M., Levermore, P., Hazell, J., Gupta, A. (2023). Open-sourcing highly capable foundation models: An evaluation of risks, benefits, and alternative methods for pursuing open-source objectives. arXiv preprint arXiv:2311.09227.

[10] Kapoor, S., Bommasani, R., Klyman, K., Longpre, S., Ramaswami, A., Cihon, P., Hopkins, A., Bankston, K., Biderman, S., Bogen, M., Chowdhury, R., Engler, A., Henderson, P., Jernite, Y., Lazar, S., Maffulli, S., Nelson, A., Pineau, J., Skowron, A., Song, D., Storchan, V., Zhang, D., Ho, D., Liang, P., Narayanan, A. (2024). On the Societal Impact of Open Foundation Models. https://crfm.stanford.edu/open-fms/paper.pdf

[11] Creative Commons, EleutherAI, GitHub, HuggingFace, LAION, & Open Future. (2023). Supporting Open Source and Open Science in the EU AI Act. https://huggingface.co/blog/assets/eu_ai_act_oss/supporting_OS_in_the_AIAct.pdf.

[12] Wladawsky-Berger, I. (2023). Are Open AI Models Safe? Linux Foundation Blog. https://www.linuxfoundation.org/blog/are-open-ai-models-safe.

[13] Raji, I & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. Conference on Artificial Intelligence, Ethics, and Society.

[14] Pipatanakul, K., Jirabovonvisut, P., Manakul, P., Sripaisarnmongkol, S., Patomwong, R., Chokchainant, P., & Tharnpipitchai, K. (2023). Typhoon: Thai Large Language Models. arXiv preprint arXiv:2312.13951.

[15] Yang, X., Wang, X., Zhang, Q., Petzold, L., Wang, W. Y., Zhao, X., & Lin, D. (2023). Shadow alignment: The ease of subverting safely-aligned language models. arXiv preprint arXiv:2310.02949.

[16] Kirchenbauer, J., Geiping, J., Wen, Y., Katz, J., Miers, I., Goldstein, T. (2023). A Watermark for Large Language Models. In Proceedings of the 40th International Conference on Machine Learning, PMLR 202:17061-17084.

[17] Han, T., Adams, L. C., Papaioannou, J. M., Grundmann, P., Oberhauser, T., Löser, A., Truhn, D., Bressem, K.K.& Bressem, K. K. (2023). MedAlpaca--An Open-Source Collection of Medical Conversational AI Models and Training Data. arXiv preprint arXiv:2304.08247.

[18] Musser, M. (2023). A Cost Analysis of Generative Language Models and Influence Operations. http://arxiv.org/abs/2308. 03740. arXiv:2308.03740 [cs].

[19] Menczer, F., Crandall, D., Ahn, Y. Y., & Kapadia, A. (2023). Addressing the harms of AI-generated inauthentic content. Nature Machine Intelligence, 5(7), 679-680. https://doi.org/10.1038/s42256-023-00690-w

[20] Lakatos, S. (2023). A Revealing Picture: AI-Generated 'Undressing' Images Move from Niche Pornography Discussion Forums to a Scaled and Monetized Online Business. Technical report. URL https://publicassets.graphika.com/reports/graphikareport-a-revealing-picture.pdf.

[21] Thiel, D., Stroebel, M., and Portnoff, R. (2023). Generative ML and CSAM: Implications and Mitigations. DOI: 10.25740/jv206yg3793. URL https://purl.stanford.edu/jv206yg3793.

[22] Castelvecchi, D. (2016). Can we open the black box of AI?. Nature News, 538(7623), 20.

[23] McIntosh, T. R., Susnjak, T., Liu, T., Watters, P., & Halgamuge, M. N. (2024). Inadequacies of large language model benchmarks in the era of generative artificial intelligence. arXiv preprint arXiv:2402.09880.

[24] Stokel-Walker, C. (2023). Critics denounce a lack of transparency around GPT-4's tech. Fast Company. https://www.fastcompany.com/90866190/critics-denounce-a-lack-of-transparency-around-gpt-4s-tech.

[25] Mills, A. (2023). FTC Investigation of ChatGPT Aims at AI's Inherent Challenges. Bloomberg Law. https://news.bloomberglaw.com/us-law-week/ftc-investigation-of-chatgpt-aims-at-ais-inherent-challenges.

[26] Ding, J., Akiki, C., Jernite, Y., Steele, A. L., & Popo, T. (2023). Towards Openness Beyond Open Access: User Journeys through 3 Open AI Collaboratives (arXiv:2301.08488). arXiv. https://doi.org/10.48550/arXiv.2301.08488.

[27] Widder, D.G., West, S., & Whittaker, M. (2023). Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI (SSRN Scholarly Paper 4543807). https://papers.ssrn.com/abstract=4543807.

[28] Liesenfeld, A., Lopez, A., & Dingemanse, M. (2023). Opening up ChatGPT: Tracking openness, transparency, and accountability in instruction-tuned text generators. Proceedings of the 5th International Conference on Conversational User Interfaces, 1–6. https://doi.org/10.1145/3571884.3604316.

[29] Nolan, M. (2023). Llama and ChatGPT Are Not Open-Source—IEEE Spectrum. IEEE Spectrum. https://spectrum.ieee.org/open-source-llm-not-open.

[30] Maffulli, S. (2023). Meta's LLaMa 2 license is not Open Source. https://blog.opensource.org/metas-llama-2-license-is-not-open-source/

[31] Open Knowledge Foundation. (2016). What is Open Data? Open Data Handbook. https://opendatahandbook.org/guide/en/what-is-open-data/ (accessed Feb. 13, 2024).

[32] Birhane, A., Steed, R., Ojewale, V., Vecchione, B., & Raji, I. D. (2024). AI auditing: The broken bus on the road to AI accountability. arXiv preprint arXiv:2401.14462.

[33] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), pp.1-42.

[34] Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P. (2020). Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 33–44. https://doi.org/10.1145/3351095.3372873

[35] Richardson, B., & Gilbert, J. E. (2021). A framework for fairness: a systematic review of existing fair AI solutions. arXiv preprint arXiv:2112.05700.

[36] Arya, V., Bellamy, R.K., Chen, P.Y., Dhurandhar, A., Hind, M., Hoffman, S.C., Houde, S., Liao, Q.V., Luss, R., Mojsilović, A. and Mourad, S. (2019). One explanation does not fit all: A toolkit and taxonomy of ai explainability techniques. arXiv preprint arXiv:1909.03012.

[37] Bird, S., Dudík, M., Edgar, R., Horn, B., Lutz, R., Milan, V., Sameki, M., Wallach, H. and Walker, K. (2020). Fairlearn: A toolkit for assessing and improving fairness in AI. Microsoft, Tech. Rep. MSR-TR-2020-32.

[38] Mökander, J., Schuett, J., Kirk, H.R., Floridi, L. (2023). Auditing large language models: A three-layered approach. AI Ethics. https://doi.org/10.1007/s43681-023-00289-2

[39] Schuett, J., Reuel, AK. & Carlier, A. (2024). How to design an AI ethics board. AI Ethics. https://doi.org/10.1007/s43681-023-00409-y

[40] Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. Proceedings of the conference on fairness, accountability, and transparency, 279-288.

[41] White, M. (2023). Shaping the Future of Generative AI with Open Source and Open Science. https://matthewdwhite.medium.com/shaping-the-future-of-generative-ai-with-open-source-and-open-science-2854005f3da5 (accessed Dec. 08, 2023)

[42] White, M. Responsible AI Framework. (2023) https://matthewdwhite.medium.com/matt-whites-responsible-ai-framework-f0385851badf (accessed Dec. 12, 2023)

[43] NTIA, Office of Public Affairs. (2023). NTIA Kicks Off Public Engagement on Executive Order AI Work. URL https://www.ntia.gov/press-release/2023/ntia-kicks-public-engagement-executive-order-ai-work.

[44] Scao, T. L., Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., Muennighoff, N., … Wolf, T. (2023). BLOOM: A 176B-Parameter Open-Access Multilingual Language Model (arXiv:2211.05100). arXiv. https://doi.org/10.48550/arXiv.2211.05100.

[45] Patel, D., & Ahmad, A. (2023). Google 'We Have No Moat, And Neither Does OpenAI'. Google 'We Have No Moat, And Neither Does OpenAI'. https://www.semianalysis.com/p/google-we-have-no-moat-and-neither.

[46] Phang, J., Bradley, H., Gao, L., Castricato, L., & Biderman, S. (2022). EleutherAI: Going Beyond "Open Science" to "Science in the Open". arXiv preprint arXiv:2210.06413.

[47] Sonnenburg, S., Braun, M. L., Cheng, S. O., Bengio, S., Bottou, L., Holmes, G., LeCun, Y., Müller, K. R., Pereira, F., Rasmussen, C. E., Rätsch, G., Schölkopf, B., Smola, A., Vincent, P., Weston, J., & Williamson, R. C. (2007). The Need for Open Source Software in Machine Learning. Journal of Machine Learning Research, 8, 2443–2466.

[48] Langenkamp, M., & Yue, D. (2022). How Open Source Machine Learning Shapes AI. AI Ethics & Society (AIES). https://doi.org/10.1145/3514094.3534167.

[49] Open Source Initiative. (2023), OSI-approved Licenses. https://opensource.org/licenses/ (accessed Dec. 08, 2023)

[50] Kerner, Sean M. (2023). Linux Foundation advances open source vision with Generative AI Commons. https://venturebeat.com/ai/linux-foundation-advances-open-source-vision-with-generative-ai-commons/ (accessed Dec. 29, 2023)

[51] European Parliament. (2023). EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

[52] Chesbrough, H. (2023). Measuring the Economic Value of Open Source. Technical Report. Linux Foundation, San Francisco, CA, USA. https://www.linuxfoundation.org/research/measuring-economic-value-of-os.

[53] Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.

[54] Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I.D. and Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19). Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[55] Data & Trust Alliance. (2023). Data Provenance Standards. https://dataandtrustalliance.org/our-initiatives/data-provenance-standards#review (accessed: Dec. 02, 2023)

[56] Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... & Yarkoni, T. (2015). Promoting an open research culture. Science, 348(6242), 1422-1425.

[57] Pineau, J., Vincent-Lamarre, P., Sinha, K., Larivière, V., Beygelzimer, A., d'Alché-Buc, F., ... & Larochelle, H. (2021). Improving Reproducibility in Machine Learning Research. Journal of Machine Learning Research, 22(107), 1-19

[58] Muñoz Ferrandis, C. (2022). OpenRAIL: Towards open and responsible AI licensing frameworks. https://huggingface.co/blog/open_rail.

[59] Open Source Initiative. (2023). Defining Open Source AI. URL
https://opensource.org/deepdive/ (accessed: Dec. 13, 2023.)

A Acknowledgements

The following members of the Generative AI Commons were instrumental in reviewing and improving upon the MOF, and their time and contributions are greatly appreciated.

- Stella Biderman, EleutherAI
- Justin Colannino, Open Source Initiative
- Stefano Maffulli, Open Source Initiative
- Ke Ding, Intel
- Ann Mary Roy, HP
- Anni Lai, Futurewei
- Ofer Hermoni, Generative AI Commons
- Nick Chase, Cloud Geometry
- Saurabh Tangri, Intel

B About The Generative AI Commons

The Generative AI Commons is a community-driven initiative at the Linux Foundation's AI & Data Foundation. It is a vendor neutral forum and open participation initiative focused on advancing principles of open science and open source in generative AI. The Generative AI Commons is dedicated to fostering the democratization, advancement and adoption of efficient, secure, reliable, and ethical Generative AI open source innovations through neutral governance, open and transparent collaboration and education.

More about the Generative AI Commons, as well as details and links to join the community can be found at https://genaicommons.org.

C About The LF AI & Data Foundation

The LF AI & Data Foundation is a global not-for-profit foundation that hosts critical components of the global AI & data technology infrastructure at the Linux Foundation.

It brings together the world's top developers, end users, and vendors to identify and contribute to the projects and initiatives that address industry challenges for the benefit of all participants.

More about the LF AI & Data Foundation can be found at https://lfaidata.foundation/