

Fostering Contractual Pathways for Responsible AI Data and Model Sharing for Generative AI and Other AI Applications

**GPAI I&C WG: Protection Innovation,
Intellectual Property (IP) Project**

November 2023



GPAI |

THE GLOBAL PARTNERSHIP
ON ARTIFICIAL INTELLIGENCE

This report was developed by Experts and Specialists involved in the Global Partnership on Artificial Intelligence's project on Protecting Innovation, Intellectual Property (IP). The report reflects the personal opinions of the GPAI Experts and External Experts involved and does not necessarily reflect the views of the Experts' organizations, GPAI, or GPAI Members. GPAI is a separate entity from the OECD and accordingly, the opinions expressed and arguments employed therein do not reflect the views of the OECD or its Members.

Acknowledgements

This report was developed in the context of the *Protecting Innovation, Intellectual Property (IP)* Project, with the steering of the Project Co-Leads and the guidance of the Project Advisory Group, supported by the GPAI Innovation & Commercialization Working Group (I&C WG). The GPAI I&C WG agreed to declassify this report and make it publicly available.

AI Data and Model Sharing Initiative Co-Leads: Lee J. Tiedrich*, Duke University; Josef Drexl*, Max Planck Institute for Innovation and Competition

GPAI IP Project Advisory Group Co-Leads: Lee J. Tiedrich*, Duke University; Yann Dietrich†, Atos/Eviden

The report was written by: Lee J. Tiedrich*, Duke University; Josef Drexl*, Max Planck Institute for Innovation and Competition; Alban Avdulla‡, Duke University, Research Assistant under the direction of Lee Tiedrich; Daria Kim‡, Max Planck Institute for Innovation and Competition, under the direction of Josef Drexl.

GPAI would like to acknowledge the tireless efforts of colleagues at the French National Institute for Research in Digital Science and Technology (Inria) and GPAI's I&C WG. We are grateful, in particular, for the support of the I&C WG Co-Chairs, Françoise Soulié and Laurence Liew, as well as Kaitlyn Bove, GPAI Project Manager at Inria.

* Expert

** Observer

† Invited Specialist

‡ Contracted Parties by the ESCs to contribute to projects

Citation

GPAI 2023. Fostering Contractual Pathways for Responsible AI Data and Model Sharing for Generative AI and Other AI Applications, Report, November 2023, Global Partnership on AI.



Table of Contents

1. AI Data and Model Sharing Initiative and Report Co-Leads	3
2. Executive Summary	4
3. Introduction	10
4. Demand for Standardized Contractual Terms Remains Strong but Efforts Are Still Relatively Nascent.	13
4.1 Summary of Certain Ongoing Contractual Initiatives to Advance AI Data and Model Sharing and Related Safety Considerations	14
4.1.1 Overview of Linux Foundation Community Data License Agreement 2.0	14
4.1.2 Overview of the Responsible AI Licenses	15
4.1.3 Overview of the Creative Commons License.....	15
5. Contracts need to be complemented by appropriate technical tools, business codes of conduct, education, and laws.	16
6. Developing Common Contractual Definitions Can Help Foster the Development of Standardized Contractual Terms	17
7. Evolving Legal Landscape Raises New Challenges and Underscores Potential Benefits of Standard Contractual Terms	18
7.1 Standardized contract terms can help address bargaining power inequities	18
7.2 Standardized contract terms can help address liability and compliance issues.....	19
7.3 Standard contract terms can help address the allocation of rights in the face of legal uncertainty	22
8. Standardized contractual terms can potentially help address practices involving the ingestion of publicly available data and code	23
9. The Importance of Multi-stakeholder Input and International Coordination.....	24
10. The IP Project Advisory Group’s Next Steps.....	24
APPENDIX 1	25
Annex 1: The Concept note.....	25
Annex 2: Workshop program.....	27
Annex 3: Discussion Outline.....	28
Session 1: The current status of the standardization of licenses for AI data and models: challenges and prospects	28
Session 2: The role of potential IP and trade secrets protection in the context of licensing AI data and models	29
Session 3: AI liability as a potential obstacle to licensing?	30
APPENDIX 2	32
Annex 1: Workshop Program	32
Annex 2: Discussion Outline and Materials to Review in Advance of Workshop.....	35
Annex 3: Background Note on GPAI AI Advisory Committee’s Work.....	39
Appendix 3: When AI generates work, standard contractual terms can help generate value and clarity	41



1. AI Data and Model Sharing Initiative and Report Co-Leads

The AI Data and Model Sharing Initiative and this report are co-led by Lee Tiedrich and Josef Drexl:

Lee J. Tiedrich is a widely recognized leader in artificial intelligence, data, and emerging technology. She is a member of both the OECD and Global Partnership on AI (GPAI) AI expert groups and co-chairs both the GPAI AI Intellectual Property Advisory Committee and the GPAI Responsible AI Strategy for the Environment Committee (RAISE). She holds three appointments at Duke University, including Distinguished Faculty Fellow in Law and Responsible Technology, Executive in Residence, and Responsible Technology Scholar in AI Health. Previously, she was a partner at the global law firm Covington & Burling LLP and an adjunct faculty member at the University of Pennsylvania Law School. A frequent speaker at leading institutions, she has held leadership positions with the American Bar Association and has served as a peer reviewer for Oxford University Press. She is a co-author of *The Law of Artificial Intelligence* (West Academic forthcoming, 2024). She served on the Biden Campaign Policy Committee and is registered to practice before the US Patent and Trademark Office.

Josef Drexl is Director of the Max Planck Institute for Innovation and Competition (Munich), Honorary Professor at the University of Munich and Member of the Bavarian Academy of Science. He was the founding Chair of the Academic Society for Competition Law (ASCOLA) from 2003 to 2013, and he is a Vice-President of the Association Internationale de Droit Economique (AIDE). Josef Drexl is an expert in competition law, intellectual property law and consumer protection law. More recent work focuses on the IP, competition, consumer and data protection law issues of the new digital economy in times of the Internet of Things and Artificial Intelligence. He is a member of the Data Governance Working Group of the Global Partnership for Artificial Intelligence (GPAI).



2. Executive Summary

The Intellectual Property (“IP”) Advisory Committee of the Innovation and Commercialization Working Group (the “Committee”) continued its work to provide a platform to help stakeholders develop standardized contractual terms to facilitate artificial intelligence (“AI”) data and model sharing. The Committee hosted two hybrid multi-stakeholder workshops, one at the Max Planck Institute for Innovation and Competition (the “MPI”) in Munich, Germany, and another at Duke University (“Duke”) in Washington, D.C, United States. The workshop concept notes and agendas are included in the Appendices. The workshops, which were undertaken in collaboration with MPI and Duke, built upon the Committee’s [2022 report](#), entitled, Preliminary Report on Data and AI Model Licensing (2022) (the “2022 Report”).

The premise of the Committee’s work is that many organizations want to share data and AI models voluntarily and responsibly, but find it challenging to do so, due to a lack of regulatory harmonization and clarity, insufficient technical tools, insufficient contractual tools and codes of conduct, and difficulties valuing the data and AI models. The 2022 report explains these challenges in more detail. The Committee’s work this year has focused on expanding pathways for organizations to develop contractual tools that can help address these challenges. Contractual terms are not necessarily the sole solution, but they can have an important role. The Committee’s work helps inform efforts to develop standard contractual terms and provides a platform for multi-stakeholder collaboration, an essential component for developing contractual terms intended for broad adoption.

The need for responsible AI data and model sharing has escalated with the rise of generative AI and other AI applications. At the [G7 Hiroshima Summit](#), G7 leaders emphasized the need to “[immediately take stock of the opportunities and challenges of generative AI](#).” They also established the [Hiroshima AI process](#), a G7 working group, to [collaborate with the OECD and GPAI](#) on addressing the pressing issues of generative AI, such as governance, safeguard of IP rights, including copyrights, promotion of transparency, and responsible uses of AI.

In a [report](#) prepared for the G7 Hiroshima AI process, the OECD identified intellectual property infringement as the second greatest risk presented by generative AI to achieving national and regional goals, according to G7 leaders. In a September statement, the G7 Digital & Tech Ministers also reaffirmed the importance of addressing intellectual property infringement and privacy threats. The Hiroshima AI Process has also resulted in International Guiding Principles and an International Code of Conduct for organizations developing advanced AI systems.

The Committee’s work advances the G7 Hiroshima AI process by creating pathways to help address certain intellectual property and other issues presented by generative AI and other applications through contracts and possibly other means. As an example, developing standard contract terms can potentially provide mutually acceptable, and more efficient, alternatives to intellectual property infringement. Standard contract terms also may help address privacy compliance when personal information is involved. The potential also exists



to help address other concerns with contracts, such as protecting a person’s likeness or publicity rights.

The evolving legal landscape also heightens the need for the Committee’s work. For example, AI continues to generate valuable works. Standardized [contractual terms can help address at least some legal uncertainties about the rights to such works](#). Additionally, some legal developments, such as the EU Data Act, aim to develop contractual terms that address inequalities of bargaining power in data-sharing practices. Having standard contractual terms may help achieve this objective, too. Standard contract terms also might help streamline AI public procurement.

Other jurisdictions are also considering the role of contracts or are undertaking initiatives where contracts can potentially have an important role in operationalizing responsible generative AI and other AI applications. For instance, the US Copyright Office released a Notice of Inquiry on AI and copyright that seeks comments on a range of issues, including *how contracts and technical tools might address rights to training data* and other developments created using generative AI. The recent [US AI Executive Order](#) contemplates more developments in this area. Canada has also launched a consultation on generative AI and copyright.

As highlighted by the [UK AI Safety Summit](#), countries and organizations, in addition to the G7, are focusing on AI safety, including for frontier AI systems. This raises questions about how AI models might be contractually licensed or made available – including on an open-source basis – in a manner that includes sufficient guardrails to prevent unethical, unsafe, and nefarious uses. It also raises questions about how to apportion responsibility and liability. Standard contract terms potentially could have a role in addressing these safety and related concerns, particularly when coupled with appropriate methods for enforcing contracts, as well as business codes or conduct, and technical tools. Open source and open innovation remain important for fostering innovation and competition.

1. Workshop Key Takeaways:

The workshops included a broad range of stakeholders spanning many geographic regions, disciplines (such as lawyers, economists, engineers, policy experts, and business representatives), and perspectives (such as policymakers, civil society, international organizations, academia, and industry). The following are key takeaways:

- Demand for Standardized Contractual Terms Remains Strong, but Efforts Are Still Relatively Nascent. The workshops confirmed that the demand for voluntary AI data and model remains strong, with the rise of generative AI and other AI applications. The demand for standardized license terms remains strong, too. These needs extend to the research community as well as commercial organizations, governments, and other stakeholders. While work has advanced since the 2022 Report, there still are no broadly embraced contractual terms for AI data or model sharing. Some current initiatives for developing standardized contractual terms vary in approach. For example, the Linux Foundation has published very simple streamlined data sharing

terms. In contrast, the Responsible AI License (“[RAIL](#)”) [community](#) has developed more detailed form agreements that seek to address AI ethical concerns in the agreements. Creative Commons licenses continue to be used by many for data sharing. Creative Commons has [explained](#) some limitations of this approach. [UNCITRAL](#) and the [Open Knowledge Foundation](#) have data licensing initiatives, too. The EU has introduced some [AI public procurement contract clauses](#). Some workshop participants expressed the view that there may be a need for multiple standard contract forms, given the many different AI use cases and applications. However, some participants commented that the universe of standardized contractual terms should not be too expansive, as this could potentially undermine the goal of standardization. Stakeholders should factor in these considerations, as they seek to develop contractual terms.

- Contracts Potentially Can Help Advance AI Safety. The Committee recognizes the mounting safety concerns of openly sharing AI models (such as through open-source licensing) and the importance of developing mechanisms that effectively prevent bad actors from using open AI models for unethical, unsafe, or nefarious purposes. The Committee believes that the community should continue to consider how contracts (and corresponding contract enforcement mechanisms, business codes of conduct and technical tools) can help contribute to supporting these safety imperatives, without unnecessarily hindering innovation.
- Contracts need to be complemented or supported by appropriate technical tools, business codes of conduct, education, and laws. As noted above, standard contractual terms will not likely be the sole solution. Appropriate technical tools, business codes of conduct, education, and laws are needed to enhance their effectiveness. Compliance with applicable privacy, intellectual property, and other laws remains paramount. Workshop participants identified several technical tools that could potentially complement standard contractual terms to facilitate responsible AI data and model sharing in a legally compliant and safe manner. These include i) watermarking or similar techniques to identify AI generated content, ii) tools to deidentify personal data and other privacy enhancing technologies (PETS), iii) data cards and other techniques that can track the provenance, lineage, and other information about the data, including information denoting whether it is copyrighted and its source, iv) AI model cards and other techniques that provide similar tracking functions for models, and v) the [W3C Text and Data Mining Reservation Protocol and similar techniques, which also may include](#) machine readable licenses. Many participants agreed that technical tools could be referenced in contracts for AI data and model sharing, making the process more streamlined. The Committee encourages stakeholders to continue to develop these types of tools and to consider how they can help advance contracting efforts. It may be helpful to draw lessons from cybersecurity and other efforts, such as developing [Software Bill of Materials](#). The US Copyright Office has solicited comments about the availability of certain technical tools. In a [joint statement issued by twelve Data Protection Authorities](#), the role of technical tools and education was highlighted in connection with addressing data scraping. Meanwhile, the UK has been working on developing appropriate [Codes of Practices/Codes of Conduct](#) to facilitate data sharing,

and the [OECD](#) has undertaken important work on business codes of conduct, too. The Hiroshima AI Process has also resulted in International Guiding Principles and an International Code of Conduct for organizations developing advanced AI systems.

- Developing Common Contractual Definitions Can Help Foster the Development of Standardized Contractual Terms. Building upon points raised in the 2022 Report, many workshop participants agreed that having standard contractual definitions could advance efforts to formulate standard contractual clauses for AI data and model sharing. The definitions should take into consideration relevant sources, such as evolving AI laws and policies, the [OECD Framework for the Classification of AI Systems](#), and potentially procurement rules. This will help ensure that contractual practices align with laws and policies.
- Evolving Legal Landscape Raises New Challenges and Underscores Potential Benefits of Standard Contractual Terms. Since the 2022 Report, several legal and policy developments, including the progress on the EU Data Act discussed above, have underscored the benefits of developing standard contractual terms for AI data and model sharing. Standard contract terms may also help advance government policies for addressing responsible AI through public procurement. Here are a couple of other examples of how contracts can potentially help advance policy goals or help navigate legal uncertainties:
 - Standardized contract terms can help address liability and compliance issues. AI applications often have complex value chains in which many entities contribute to their development. For instance, some entities may supply data or AI models upstream that are used by other entities further downstream in connection with AI application development. Oftentimes, the upstream data and AI model suppliers do not have visibility or cannot foresee exactly how their data and models will be used by others downstream. This can make it more challenging for them to mitigate potential safety and other harms and liabilities that can potentially arise downstream from the use of their data or models supplied upstream. Policy makers and other stakeholders continue to consider mechanisms for allocating responsibility and liability in complex AI value chains. Workshop participants expressed interest in exploring how contract terms and technical standards might support or otherwise complement legal developments. They also raised questions about how legal developments might impact customary contracting practices of disclaiming warranties and limiting liability using “as is and where is” contract terms, in at least some scenarios.
 - Standard contract terms can help address allocation of rights in the face of legal uncertainty.

The rapid development of generative AI and other AI applications has heightened the IP debate and raised more IP questions. For instance, questions about authorship and copyright ownership of AI-generated and AI-aided outputs are attracting significant legal and policy attention. The same is true for the allocation of responsibility and liability for AI-generated works that infringe upon the rights of others, and the scope of fair use and text and data mining (TDM) exceptions and the extent to which “likeness” and publicity rights are protected. Similar types of uncertainties also impact the patent landscape. As the legal landscape continues to evolve, workshop participants generally agreed that contracts can potentially help parties establish at least some more certainty to help allocate rights and guide their activities. The Committee encourages the community to continue to explore ways that contracts can responsibly and fairly advance these efforts.

- Standard contract terms can potentially help with trade secret protection. Trade secrets garnered significant attention at the workshops. Insights from an EU-mandated study¹ that surveyed companies’ perceptions regarding the role and value of trade secrets protection were presented and discussed. Specifically, concerns were shared that once a party starts broadly transacting data presumed to be protected as trade secrets, uncertainty arises about whether that party still maintains the value of trade secrets and is in the position to ensure reasonable measures that would maintain the trade secrets protection. At the same time, if recipients of trade secrets are obliged to uphold confidentiality and ensure the appropriate level of data security, this can have a restrictive effect on data-sharing practices. The Committee encourages the community to consider how contracts might be drafted that provide for data sharing while reasonably preserving trade secret protections, when desired.
- Standardized Contractual Terms Potentially Can Help Address Practices Involving the Ingestion of Publicly Available Data and Code. Many AI applications require access to significant amounts of AI data, including for training, testing and validation. Stakeholders are aggregating AI data through a variety of means, including scraping or ingesting data from third party websites and social media properties. In addition, organizations increasingly are scraping open source and other publicly available computer code in connection with the development of AI models. These practices have skyrocketed, along with the meteoric rise of generative AI. Not surprisingly, litigation and enforcement has ensued, and policy proposals have emerged addressing the scraping or ingestion of at least some data. Data and code scraping and ingestion practices present intellectual property, privacy, consumer protection, “likeness,” and other legal issues. Relevant laws also vary among jurisdictions. As highlighted in a recent WIPO conversation about Generative AI and IP, harmonizing laws across jurisdictions remains challenging. While further study and consideration

¹ Radauer A, Bader M, Aplin T et al., *Study on the Legal Protection of Trade Secrets in the Context of the Data Economy: Final report* (Publications Office of the European Union 2022) <https://data.europa.eu/doi/10.2826/021443>.

is needed, several workshop participants agreed that standard contractual terms – in combination with business codes of conduct, technical tools, education, and laws – could potentially help foster more responsibility and clarity with respect to the ingestion and use of publicly available data and code. The Committee believes that these efforts merit further study and consideration, given the prevalence of these practices and the high cost and other challenges of litigation and enforcement. Any resulting solution must comply with applicable laws and be fair to all stakeholders, including small and medium-sized enterprises (SMEs) and historically under-represented groups.

- Multistakeholder input is essential. Workshop participants generally agreed that multi-stakeholder participation is essential for drafting standard contractual terms that will be broadly adopted.
- International Regulatory Harmonization. Workshop participants generally agreed that increasing international regulatory harmonization can facilitate responsible and voluntary AI data and model sharing.

2. Committee Next Steps:

Considering the feedback and valuable recommendations of many participants, the Committee plans to continue its work during 2024. Participants showed great interest and acknowledged the importance of multi-stakeholder collaboration in trying to address these issues.

The Committee plans to launch an AI Contract Terms Incubator (AI CTI), with the goal of providing stakeholders with a forum to share ideas and get feedback from each other through virtual meetings and possibly at least one hybrid workshop. The AI CTI is still in the planning phase. The Committee would like the AI CTI to enhance opportunities for stakeholders to share draft contract terms or clauses with other AI CTI participants and solicit their comments. Furthermore, the AI CTI may serve as a forum for soliciting feedback on ideas and approaches that potentially could be used by participants to develop standard contractual terms. The Committee plans to make the incubator open to all interested participants worldwide, including academia, civil society, government, and industry. Parties interested in participating in the AI CTI should contact Kaitlyn Bove (kaitlyn.bove@inria.fr).



3. Introduction

Advancements in artificial intelligence (AI), along with AI-driven progress across various sectors, require access to data. Data serves as input for numerous purposes and activities within the AI value chain, including the training, testing, and validation of machine learning (ML) models and systems. The voluntary sharing of data assumes a pivotal role in meeting the demand for data.

The need to support and facilitate the voluntary sharing of AI data² and ML models has been acknowledged by stakeholders, including companies, governments, academia, inter-governmental, and non-profit organizations. However, challenges in AI data and model sharing persist due to various factors, including the lack of incentives or motivation among data holders, caused by and further exacerbated by issues such as legal uncertainties and economic concerns, including those related to competitive advantage. Additionally, transaction costs play a role, along with the fragmentation of data and the absence of interoperability at both technical and semantic levels. These problems can be compounded by the potential for contractual imbalances and unequal distribution of bargaining power.

At the same time, a consensus is emerging that standardized contract terms can help surmount some of these challenges by reducing transaction costs, fostering more legal certainty, and advancing other related goals. Several initiatives are underway to develop standard contractual terms that would address the unique aspects of data and AI models, particularly by expanding tools for the voluntary sharing of AI models, whether through open-source arrangements or other terms.³ Most initiatives have developed organically through a bottom-up approach, starting with identifying specific needs within a community before proposing any standardized solutions. These still relatively nascent efforts encounter various technical, economic, legal, institutional, and business challenges persisting on the way to achieving optimal levels of responsible and efficient voluntary AI data and model sharing.

The Committee recognizes the importance of fostering more responsible and efficient sharing of AI data and models to unlock the promise of responsible AI. For this reason, the Committee launched a project in 2021 aiming to assess and support ongoing initiatives to develop standard contract terms for AI data and model sharing. The initial phase of the project focused on identifying the challenges confronting these efforts and exploring potential solutions. This was accomplished through conducting interviews with various stakeholders and consultations among members of the GPAI Multi-stakeholder Expert Group (MEG). The outcome of this endeavor was the publication of the 2022 Report, summarizing the Committee's findings up to September 2022. The report highlights the legal, technical, and business challenges pertaining to the sharing of AI data and models. It also identifies areas requiring further research and suggests potential future actions. One of these actions involves fostering an inclusive ecosystem where a diverse range of stakeholders can access relevant information and expertise and collaborate to advance the development of standardized contract terms for

² By 'AI data', this report refers collectively to datasets used for the development of AI models and applications, including training, testing, and validation datasets.

³ For an overview, see [2022 Preliminary Report on Data and AI Model Licensing](#).

sharing AI data and models. In pursuit of this objective, the Committee planned and carried out two multi-stakeholder workshops co-organized and hosted by the [Max Planck Institute for Innovation and Competition](#) and [Duke University](#). These workshops took place in April and June 2023, respectively.⁴ The workshop concept notes and agendas are included in the Appendices.

The need for responsible AI data and model sharing has escalated with the rise of generative AI and other AI applications. At the [G7 Hiroshima Summit](#), G7 leaders emphasized the need to “[immediately take stock of the opportunities and challenges of generative AI](#).” They also established the [Hiroshima AI process](#), a G7 working group, to [collaborate with the OECD and GPAI](#) on addressing the pressing issues of generative AI, such as governance, safeguard of IP rights, including copyrights, promotion of transparency, and responsible uses of AI.

In a [report](#) prepared for the G7 Hiroshima AI process, the OECD identified intellectual property infringement as the second greatest risk presented by generative AI to achieving national and regional goals, according to G7 leaders. In a September statement, the G7 Digital & Tech Ministers also reaffirmed the importance of addressing intellectual property infringement and privacy threats. The Hiroshima AI Process has also resulted in International Guiding Principles and an International Code of Conduct for organizations developing advanced AI systems.

The Committee’s work advances the G7 Hiroshima AI process by creating pathways to help address certain intellectual property and other issues presented by generative AI and other applications through contracts and possibly other means. As an example, developing standard contract terms can potentially provide mutually acceptable, and more efficient, alternatives to intellectual property infringement. Standard contract terms and codes of conducts also may help address privacy compliance when personal information is involved. The potential also exists to help address other concerns with contracts, such as protecting a person’s likeness or publicity rights.

The evolving legal landscape also heightens the need for the Committee’s work. For example, AI continues to generate valuable works. Standardized [contractual terms can help address at least some legal uncertainties about the rights to such works](#). Additionally, some legal developments, such as the EU Data Act, aim to develop contractual terms that address inequalities of bargaining power in data-sharing practices. Having standard contractual terms may help achieve this objective, too. Standard contract terms also might help streamline AI public procurement.

Other jurisdictions are also considering the role of contracts or are undertaking initiatives where contracts can potentially have an important role in operationalizing responsible generative AI and other AI applications. For instance, the US Copyright Office released a [Notice of Inquiry on AI and copyright](#) that seeks comments on a range of issues, including

⁴ The workshop ‘Exploring Pathways to the Standardization of Licenses for Data and Machine Learning Models’ was held at the Max Planck Institute for Innovation and Competition in Munich on 28-29 April 2023. The workshop ... was held at Duke University on ... 2023.



how contracts and technical tools might address rights to training data and other developments created using generative AI. The recent US AI Executive Order contemplates more developments in this area. [Canada](#) has also launched a consultation on generative AI and copyright.

As highlighted by the UK AI Safety Summit, many countries and organizations, in addition to the G7, are focused on AI safety, including for frontier AI applications. This raises questions about how AI models might be contractually licensed, including on an open-source basis, in a manner that includes sufficient guardrails to prevent unethical, unsafe, and nefarious uses. It also raises questions about how to apportion responsibility and liability. Standard contract terms potentially could have a role in addressing these safety and other concerns too, particularly when coupled with appropriate methods for enforcing contracts, business codes of conduct, and technical tools. Open source and open innovation remain important for fostering innovation and competition.

The workshops were conducted pursuant to [Chatham House Rules](#). This report provides a high-level summary of the workshop discussions, consistent with Chatham House Rules.

In summarizing the workshop discussions, the authors of this report did not independently verify the accuracy or completeness of statements made by workshop participants and relied on the accuracy and completeness of such statements. In some sections, the authors did include information about important developments that occurred or came to their attention after the workshops, such as the US Copyright Office's Notice of Inquiry on AI and Copyright and the similar Canadian consultation.



4. Demand for Standardized Contractual Terms Remains Strong but Efforts Are Still Relatively Nascent.

AI models are trained, tested, and validated on vast amounts of data, and recent developments, including the rise of generative AI, indicate that the [demand for data remains strong](#). As reflected in the Committee's 2022 Report, stakeholders expressed strong interest in having more mechanisms and tools for responsibly and voluntarily sharing data to advance AI, including contractual mechanisms and technical tools.

Many participants in the Committee's 2023 Workshops observed that the demand for these mechanisms remains strong, along with the demand for applicable business codes of conduct and technical tools for AI data and model sharing. Such demand appears to be fueled, at least in part, by the growth of generative AI and other AI applications. It also appears to be fueled by the potential for [contract terms to help provide organizations with more certainty](#), particularly while relevant IP and other laws continue to evolve, as explained more fully below. Several workshop participants also reported that the demand for AI data and model sharing extends to the research community, in addition to governments, commercial organizations, and other stakeholders.

Despite such demand, several workshop participants observed that the development of standard contract terms for AI data and model sharing remains relatively nascent. Although important progress has been made since last year, no standard AI data or model sharing terms have been widely adopted and embraced throughout society. Many workshop participants also expressed the view that standard contract clauses, technical tools, and business codes of conduct might have the added benefit of helping organizations comply with emerging AI laws and regulations, including the EU Artificial Intelligence Act (the "EU AI Act"). They also may help with government procurement of AI. As discussed below, workshop participants also recognize that more work needs to be done to create standardized technical tools and business codes of conduct for AI data and model sharing, in addition to standardized contract terms.

The Committee's [2022 Report](#) highlighted some ongoing efforts to develop standard contract terms, including those of [the Linux Foundation](#) and the [Responsible AI Licenses](#) community (RAIL). Workshop participants discussed these and other efforts, as well as the use of [Creative Commons](#) licenses for data sharing. Workshop participants also discussed initiatives to build data commons, including by the [French Health Data Hub](#) and [ML Commons](#).



4.1 Summary of Certain Ongoing Contractual Initiatives to Advance AI Data and Model Sharing and Related Safety Considerations

Some current initiatives for developing standardized contractual terms for AI data and model sharing vary in approach. For example, the Linux Foundation has published very simple streamlined data sharing terms. In contrast, RAIL has developed more detailed form agreements that seek to address ethical concerns raised by AI applications. [UNCITRAL](#) and the [Open Knowledge Foundation](#) have data licensing initiatives, too. The EU has introduced some [AI public procurement contract clauses](#). [Creative Commons](#) has contributed to the [dialogue](#) too, and it appears that many continue to use its licenses for AI data sharing.

Some workshop participants expressed the view that there may be a need to have multiple standard contract forms for AI data and model sharing, given the many different AI use cases and applications. The Committee currently shares this perspective. Some workshop participants commented that the universe of standardized contractual terms should not be too expansive, as this could potentially undermine the goal of standardization. Stakeholders should factor in these considerations, as they seek to develop standard contractual terms.

Finally, the Committee recognizes the mounting safety concerns of openly sharing AI models (such as through open-source licensing) and the importance of developing mechanisms that effectively prevent bad actors from using open AI models for unethical, unsafe, or nefarious purposes. The Committee believes that the community should continue to consider how contracts -- along with contract enforcement mechanisms, business codes of conduct, and technical tools -- can help contribute to supporting these safety imperatives without unnecessarily hindering innovation.

4.1.1 Overview of Linux Foundation Community Data License Agreement 2.0

[The Linux Foundation](#) has developed the [Community Data License Agreement 2.0](#), a simple permissive license designed to facilitate easy data sharing. While the Linux Foundation affirms its support of trustworthy AI, ethics and responsible use are not addressed in this license agreement in order to keep it simple and to foster greater uniformity and compatibility. This Linux license reflects the perspective that license restrictions and license customization can give rise to incompatibility, which in turn can present obstacles to data sharing. This view informs the Linux Foundation's approach to keep its standard license simple. The Linux Foundation recommends addressing AI trustworthiness outside of this data licensing agreement, including with technical tools, standards, other agreements, and/or certifications.



4.1.2 Overview of the Responsible AI Licenses

[Responsible AI Licenses](#) (RAIL) has developed several licenses, such as i) the [Responsible AI Pubs Licenses](#), which includes model and source code licenses, ii) the [Responsible AI End-User License](#), iii) the [Responsible AI Source Code License](#), and iv) the [BigScience Open RAIL-M License](#). In contrast to the Linux license discussed above, RAIL licenses include [behavioral-use clauses](#) that enable developers to permit or prohibit particular use cases. RAIL believes this is important to promote responsible downstream use.

The RAIL licenses can be slightly modified to provide some flexibility. Its approach focuses on “*imposing similar use restrictions.*” Distributing the source code may require the code provider to share a copy of the license, a model card, and explanatory documentation, which to help downstream.

4.1.3 Overview of the Creative Commons License

Workshop participants reported that many stakeholders continue to rely on Creative Commons licenses for AI data sharing. In blog posts, however, [Creative Commons](#) has emphasized that [its licenses were drafted for copyrighted works and may not address all relevant issues arising in connection with data sharing](#), such as privacy and ethical concerns. Additionally, some data may not be eligible for copyright protection.



5. Contracts need to be complemented by appropriate technical tools, business codes of conduct, education, and laws.

As noted above, standard contractual terms will not likely be the sole solution. Appropriate technical tools, business codes of conduct, education, and laws are needed to enhance their effectiveness. Compliance with applicable privacy, intellectual property, and other laws remains paramount. Workshop participants identified several technical tools that could potentially complement standard contractual terms to facilitate responsible AI data and model sharing in a legally compliant manner. These include i) watermarking and other techniques to identify AI generated content, ii) tools to deidentify personal data and other privacy enhancing technologies (PETS), iii) data cards and other techniques that can track the provenance, lineage, and other information about the data, including information denoting whether it is copyrighted and its source, iv) AI model cards and other techniques that provide similar tracking functions for models, and v) the [W3C Text and Data Mining Reservation Protocol](#) and similar techniques, which also may include machine readable licenses.

While work on technical tools continues to advance, most workshop participants acknowledged a need to increase efforts to standardize some of the emerging technical tools. For example, some workshop participants reported that widely adopted standardized data cards and AI model cards do not exist. Many participants agreed that the technical tools could be referenced in contracts for AI data and model sharing and that standardized tools might streamline the contracting process. In addition, standardized tools -- such as data cards and AI model cards -- potentially can help create mechanisms that can help allocate liability and responsibility among parties and aid with regulatory compliance and enforcement. The Committee encourages stakeholders to continue to develop these types of tools and to consider how they can help advance contracting and other efforts. Stakeholders should consider whether such tools should have features that can be customized for particular sectors, such as health data and environmental data. This work potentially may be informed by cybersecurity and other efforts, such as developing [Software Bill of Materials](#).

The Committee also is encouraged by ongoing efforts to develop business codes of conduct, including by the [OECD](#) and the [United Kingdom's Information Commissioner's Office](#). In connection with the Hiroshima AI Process, the G7 has created international guiding principles and a code of conduct for the development of advanced AI systems. Efforts to develop contract terms should factor in these important initiatives as well.



6. Developing Common Contractual Definitions Can Help Foster the Development of Standardized Contractual Terms

Building upon the 2022 Report, many workshop participants agreed that having standard contractual definitions could advance efforts to formulate standard contractual clauses for AI data and model sharing. The definitions should take into consideration relevant sources, such as evolving AI laws and policies, the [OECD Framework for the Classification of AI Systems](#), and potentially procurement rules. This will help ensure that contractual practices align with laws and policies. Definitions could be crafted to denote AI input data, AI prompts, untrained AI models, trained AI models, weights, AI outputs, and various types of processed data. It would be helpful to develop such terms with multi-stakeholder input in order to encourage broad adoption.

Standard contractual definitions could lay the foundation for broader standard contractual clauses. In addition, these definitions could help streamline efforts to negotiate bespoke AI data or model sharing agreements, when such an approach is desired. The Committee recognizes that standard contract terms will not supplant the need for bespoke terms in some contexts.

7. Evolving Legal Landscape Raises New Challenges and Underscores Potential Benefits of Standard Contractual Terms

The increasing complexity of the evolving legal environment in which AI data and model sharing practices take place poses uncertainties and challenges,⁵ as already stated in the 2022 Report. This surge in policy and legislative activity highlights the importance of addressing interoperability challenges in the field of AI. Legal uncertainty is a factor of hesitancy in sharing AI data and models, and navigating legal uncertainty becomes a challenge, especially in a cross-border context. Adding to the challenge, the relevant legal issues span IP, privacy, rights of publicity or “likeness”, consumer protection, and other areas.

The question then arises: Could standardizing contractual terms assist in managing this uncertainty and enhancing the facilitation of these transactions? Workshop participants generally agreed that contractual terms could help address at least some uncertainty, particularly when combined with business codes of conduct, education, and technical tools.

7.1 Standardized contract terms can help address bargaining power inequities

Unfair disparities in bargaining power between some companies holding data and some potential data users have been posited among the challenges for attaining an optimal level of data sharing. Concerns that such imbalances can be exploited – such as by imposing economically prohibitive conditions on the data user, in particular, SMEs – were shared by workshop participants.

In principle, when it comes to addressing inequalities in bargaining power, a number of potential contractual solutions come within the possible purview, including model contract terms, the control of unfair contract terms, and default contractual rules. The Committee does not take any views with respect to the need for, or the form of, regulation, as policy recommendations are outside the scope of the Committee’s work. The Committee is focused on fostering the development of standard contract terms that will facilitate voluntary and responsible AI data and model sharing.

The EU is seeking to proactively address the risk of exploiting discrepancies in bargaining power, with a view to fostering a data-driven economy. Concerns about contractual imbalances have been expressed in numerous policy initiatives aimed at promoting a data-driven economy in the EU and ultimately culminated in a proposal for the EU Data Act.⁶ This

⁵ According to the OECD, the past few years have witnessed a significant increase in the number of AI-related legislation worldwide. OECD, *The State of Implementation of the OECD AI Principles. Insights from National AI Policies* (OECD ...) <https://oecd.ai/en/policies>.

⁶ Art. 13 of the Proposal of the Commission of 23 February 2022 for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM(2022) 68 final (hereinafter the proposed EU Data Act). At the time of writing, the proposal for the EU Data Act has passed the first reading at the European Parliament.

legislative precedent was a topic of discussion during the workshops, particularly the one hosted by Duke. Once the EU Data Act comes into effect, some of its instruments hold relevance and potential value for the international community working on the standardized terms for AI data and model-sharing agreements, as explained below. The Committee seeks to help foster the development of standard contract terms that promote responsible AI data and model sharing.

The proposed EU Data Act envisages enhancing the availability of data for utilization in all economic sectors and in different scenarios, in accordance with the overarching principle of fairness. This concept of fairness also translates into the rules on contracts for making data available, in particular, addressing the risk of the imbalances in bargaining power being misused. On the one hand, it is acknowledged that contractual terms that are commercially more favorable to one party are normal manifestations of contractual freedom, especially in business-to-business commercial practices.⁷ On the other hand, unilaterally imposed contractual terms that ‘grossly deviate from good commercial practice in data access and use, contrary to good faith and fair dealing,’ would be deemed unfair and shall not be binding under the EU Data Act.⁸

Overall, the draft EU Data Act contains a number of mechanisms intended to ensure fairness of contractual data-sharing practices, including control of unfair contract terms,⁹ mandatory contract law regarding minimum obligations on data processing service providers,¹⁰ and model contract terms.¹¹ The latter are of particular interest for the Committee’s work. In particular, the European Commission has the mandate to propose i) non-binding model contract terms regarding data access and use, including reasonable compensation and the protection of trade secrets, and ii) non-binding standard contractual clauses for cloud computing contracts that would aid parties in ‘drafting and negotiating contracts with fair, reasonable and nondiscriminatory contractual rights and obligations’.¹² These terms are expected to reflect sector specificities, aid smaller businesses in contract negotiations, and when widely adopted, promote more equitable data access and sharing agreements.¹³ The work of the respective working parties at the European Commission is still in relatively nascent stages. The Committee can serve as a resource to working parties undertaking this work.

7.2 Standardized contract terms can help address liability and compliance issues

Nearly in every interview carried out by the Committee in 2022, participants pointed out that liability issues in the context of AI innovation pose significant uncertainty for AI data and model

The consolidated compromise text of the Proposal is available at: [https://www.europarl.europa.eu/RegData/commissions/itre/lcag/2023/07-14/ITRE_LA\(2023\)004595_EN.pdf](https://www.europarl.europa.eu/RegData/commissions/itre/lcag/2023/07-14/ITRE_LA(2023)004595_EN.pdf).

⁷ Recital 54 of the proposed EU Data Act.

⁸ Article 13(1) and (3) of the proposed EU Data Act.

⁹ Article 13 of the proposed EU Data Act.

¹⁰ Article 24(1) of the proposed EU Data Act.

¹¹ Article 34 of the proposed EU Data Act.

¹² Art 34 of the proposed EU Data Act.

¹³ Rec 83(b) of the proposed EU Data Act.

sharing. Such issues include the following two categories: i) liability for harm caused to third parties arising from AI systems or applications related to a data- or model-sharing as well as compliance with the prospective safety regulations, and ii) liability for non-compliance with the safety regulations targeting the development and deployment of AI systems. As the AI regulatory landscape continues to evolve, the introduction of technical standards and AI-specific compliance and liability rules pose considerable uncertainty in the realm of contractual sharing of AI data and models.

A peculiar aspect of AI data and model sharing is that AI data and models often function as ‘upstream’ inputs for components of AI systems. Oftentimes, the ‘downstream’ consequences and events of the AI system can hardly be predicted by the upstream parties when the contract is formed. Where data or AI models are used as input for research and development (R&D) or product development in a complex value chain, there is a potential for harm or damage to be caused or mediated by the output. It is, therefore, unsurprising that upstream parties providing data or AI models for downstream use in R&D or in a complex value chain usually seek to minimize contractual warranties and liability and offer the data and AI models on an “as-is, where-is” basis. In many cases, providing more warranties and assuming more liability can be challenging. This situation frequently arises because the downstream use and value of data and AI models can be highly contextual, and ensuring the quality of the prospective model performance can be challenging for upstream suppliers, if not impossible.

AI-related liabilities and their potential effects on AI data and model sharing were a particular focus of the Munich workshop, which was not coincidental, given several ongoing EU legislative initiatives addressing safety and liability issues in the context of AI. Such targeted initiatives include the proposal for the EU AI Act,¹⁴ the revised Product Liability Directive that will adapt EU product liability rules to new technologies (specifically AI),¹⁵ and a new AI Liability Directive relating to fault-based claims for damages that occur through AI applications.¹⁶

The upcoming EU AI Act represents a comprehensive regulation, *inter alia* regarding the quality of data used for the development of high-risk AI systems. It will require development of technical standards and implement an AI-specific compliance regime. This includes a focus on the data quality utilized for the development of high-risk AI systems.

The enactment of such legislative instruments raises questions about their implications for the contractual allocation of access and usage rights in data and AI models, particularly within a cross-border context. The discussion during the Munich workshop, which was moderated to identify major areas of legal uncertainty and explore potential solutions, delved into the

¹⁴ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (21 April 2021) COM (2021) 206 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>. At the time of writing, the proposal is awaiting the European Parliament’s position in the first reading.

¹⁵ Proposal for a Directive of the European Parliament and of the Council on liability for defective products (28 September 2022) COM(2022) 495, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0495&qid=1676546081706&from=EN>.

¹⁶ Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) (28 September 2022) COM(2022) 496, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496>.

aspects that should be considered when drafting liability clauses in such scenarios. It also sought perspectives on whether and to what extent “as-is, where-is” provisions might be deemed compliant with the upcoming liability rules.

Given the nature of these regulations, it is unsurprising that many workshop participants perceived them as causing significant uncertainty within the realm of contractual arrangements. The discussion predominantly centered on the requirements for data and data governance of high-risk AI systems, with a specific emphasis on data quality (Article 10 of the draft EU AI Act), which are anticipated to exert substantial impact on the contractual allocation of access and use rights in AI data and models. A viewpoint widely held, particularly among technical specialists, is that complying with this provision would be challenging, if not unfeasible in practice unless accompanied by explicit guidance and technical specifications. The EU AI Act contemplates the development of such technical specifications. Potentially, the obligations regarding data quality and robustness of the training process under the EU AI Act for high-risk AI systems may render the unqualified and unconditional “as-is, where-is” approach in data-sharing agreements no longer possible, in at least some circumstances. The development of technical specifications and tools potentially could inform alternative contractual approaches. For example, contracts could possibly refer to the use of, and compliance with, such technical specifications and tools. The Committee encourages the community to continue to explore these approaches.

Furthermore, concerns were expressed regarding the expansion of the scope of Article 10 of the proposed EU AI Act to apply to general-purpose AI systems. This expansion could potentially make the upstream AI data or model suppliers accountable for downstream applications that were unforeseen at the time of concluding a data or code-sharing agreement. These novelties are expected to prompt a re-evaluation of the currently employed contractual terms. Notably, according to empirical results presented at the workshop, many companies supplying AI data or models upstream experience difficulties in determining whether downstream AI systems can be classified as high-risk under the draft EU AI Act.¹⁷ The Committee would encourage the community to continue to explore how contracts can help parties responsibly and fairly allocate liability and address these issues, in a way that complies with applicable laws.

Some workshop participants expressed the view that the evolving legal landscape continues to complicate the process of developing standard contractual terms for AI data and model sharing. For instance, they explained their perspective that several important legal questions remain open, including i) how will the statutory and contractual allocation of liability interact with licenses for AI data and models; ii) how much leeway would parties have to contractually define the scope of liability, the types of harm, indemnification, and the applicable substantive law as regards liability; and iii) what are the implications of these upcoming acts for cross-border transactions for AI data and models. The question of whether data governance obligations under the upcoming EU AI Act for high-risk AI systems may eventually necessitate the inclusion of new clauses in the data- or model-sharing agreements requires further analysis.

¹⁷ Liebl A and Klein T, *AI Act: Risk Classification of AI Systems from a Practical Perspective* (2023), <https://www.appliedai.de/assets/files/AI-Act-Risk-Classification-Study-EN.pdf>.

Some legal experts expressed the view that, even if parties may no longer be able to assign liability for harm arising from a contract in a way other than prescribed by statutory law, it would still be beneficial for parties to specify in a contract their respective cooperation duty in such an event. The Committee does not express any views on how the underlying laws should be developed or enforced. However, the Committee wants to share the perspective of workshop participants for consideration by policymakers and the broader community and advance efforts to develop contract terms.

7.3 Standard contract terms can help address the allocation of rights in the face of legal uncertainty

The rapid development of generative AI and other AI applications has heightened the IP debate and raised more IP questions. For instance, questions about authorship and copyright ownership of AI-generated and AI-aided outputs are attracting significant legal and policy attention. The same is true for the allocation of responsibility and liability for AI-generated works that infringe upon the rights of others, and the scope of fair use and text and data mining (TDM) exceptions and rights of publicity and likeness. Similar types of uncertainties also impact the patent landscape. As the legal landscape continues to evolve, workshop participants generally agreed that contracts can potentially help parties establish at least some more certainty to allocate rights and guide their activities. The Committee encourages the community to continue to explore ways that contracts can responsibly and fairly advance these efforts.

Trade secrets garnered significant attention at the workshops. Insights from a recent EU-mandated study¹⁸ that surveyed companies' perceptions regarding the role and value of trade secrets protection in the digital era were presented and discussed. Specifically, concerns were shared that once a party starts broadly transacting data presumed to be protected as trade secrets, uncertainty arises about whether that party still maintains the value of trade secrets and is in the position to ensure reasonable measures that would maintain the trade secrets protection. At the same time, if recipients of trade secrets are obliged to uphold confidentiality and ensure the appropriate level of data security, this can have a restrictive effect on data-sharing practices. The Committee encourages the community to consider how contracts might be drafted that provide for data sharing while reasonably preserving trade secret protections, when desired.

¹⁸ Radauer A, Bader M, Aplin T et al., *Study on the Legal Protection of Trade Secrets in the Context of the Data Economy: Final report* (Publications Office of the European Union 2022) <https://data.europa.eu/doi/10.2826/021443>.



8. Standardized contractual terms can potentially help address practices involving the ingestion of publicly available data and code

Many AI applications require access to significant amounts of AI data, including for training, testing and validation. Stakeholders are aggregating AI data through a variety of means, including scraping or ingesting data from third party websites and social media properties (“data ingestion”). In addition, organizations increasingly are ingesting open source and other publicly available computer code in connection with the development of AI models. These practices have skyrocketed, along with the meteoric rise of generative AI.

Not surprisingly, litigation and enforcement actions involving data and code ingestion have ensued. Additionally, legislative proposals have emerged addressing at least some of these practices. These cases and proposals are complex, since data and code ingestion can raise intellectual property, privacy, consumer protection, rights of publicity, “likeness,” and other legal issues. Complicating matters, the facts underlying the cases can vary, and litigation and enforcement can be time consuming and expensive. The Committee heard from privacy and intellectual property experts who confirmed that the legal matters are complex. Relevant laws also vary among jurisdictions. As highlighted in a recent WIPO conversation about [Generative AI and IP](#), harmonizing laws across jurisdictions remains challenging.

While further study and consideration is needed, several workshop participants agreed that standard contractual terms – in combination with business codes of conduct, technical tools, education, and laws – could potentially help foster more responsibility and certainty with respect to the ingestion and use of data and code. Some jurisdictions recognize that contracts can be used to help manage these practices. The US Copyright Office has solicited comments on this topic and other issues involving generative AI and copyright. In a [joint statement](#), the twelve Data Protection Authorities emphasized the significance of technical tools and education concerning data scraping. The UK government is working on a [business code of conduct](#), and the [OECD](#) has various business codes of conduct. In connection with the Hiroshima AI Process, the G7 has created international guiding principles and a code of conduct for the development of advanced AI systems.

The Committee believes that efforts to develop contract terms for data and code ingestion merit further study and consideration. Any resulting solution must comply with applicable laws and be fair to all stakeholders, including SMEs and historically under-represented groups. This approach also holds the promise of being more efficient, particularly given the time and cost associated with litigation and enforcement. Since many parties are affected by data and code ingestion, consideration of contractual and other solutions should engage a broad range of stakeholders.



9. The Importance of Multi-stakeholder Input and International Coordination

The pursuits to develop standard contractual terms for AI data and models should seek input from multiple stakeholders. This necessity arises from the pervasive cross-sector and societal impact of AI technology. Workshop participants agreed, consistent with the outcomes of the recent G7 Summit, that multi-stakeholder participation and collaboration are critical for addressing multiple challenges related to AI data and model sharing and drafting standard contractual terms that can be broadly adopted. Workshop participants also agreed that increasing international regulatory harmonization can facilitate responsible and voluntary AI data and model sharing.

10. The IP Project Advisory Group's Next Steps

Considering the feedback and valuable recommendations of many participants, the Committee plans to continue its work during 2024. Participants showed great interest and acknowledged the importance of multi-stakeholder collaboration in trying to address these issues.

More specifically, the Committee plans to launch an AI Contract Terms Incubator (AI CTI), with the goal of providing stakeholders a forum to share ideas and get feedback from each other through virtual meetings and possibly at least one hybrid workshop. The AI CTI is still in the planning phase. The Committee would like the AI CTI to enhance opportunities for stakeholders to share draft contract terms or clauses with other AI CTI participants and solicit their comments. Furthermore, the AI CTI may serve as a forum for soliciting feedback on ideas and approaches that potentially could be used by participants to develop standard contractual terms. The Committee hopes to make the incubator open to all interested participants worldwide, including academia, civil society, government, and industry. Parties interested in participating in the AI CTI should contact Kaitlyn Bove: kaitlyn.bove@inria.fr

APPENDIX 1

‘Exploring Pathways to the Standardization of Licenses for Data and Machine Learning Models’

A workshop co-organized by the Global Partnership on Artificial Intelligence, the MaxPlanck Institute for Innovation and Competition (Munich), and Duke University

Annexes to the invitation

Annex 1: The Concept note

In recent years, it has become widely acknowledged that access to data is crucial for artificial intelligence (AI) innovation and that voluntary data sharing can help address this need. AI depends on appropriate data for many purposes, including training models, testing, and validation (collectively, ‘AI data’). Interest also has increased in expanding tools for the voluntary sharing of AI models, whether on an open-source basis or pursuant to other terms. Despite these pressing needs, society has not yet achieved optimal levels of responsible and efficient AI data and model sharing due to several legal, technical, and economic challenges. It is therefore imperative for stakeholders, including companies, governments, academia, and non-profit organizations, to work collaboratively to address these challenges with the goal of facilitating AI data and model sharing.

There is emerging consensus that standardized contract terms can help overcome some of these challenges by reducing transaction costs, fostering more legal certainty, and helping to advance other related goals. Relatively nascent efforts are underway to develop such standardized terms.

The IP Project Advisory Group of the Innovation and Commercialization (I&C) Working Group of the Global Partnership on Artificial Intelligence (GPAI) recognizes the importance of fostering more responsible and efficient AI data and model sharing in order to help unlock the promise of AI. For this reason, the GPAI IP Project Advisory Group launched a project in 2022 to assess and support ongoing initiatives to develop standard contract terms for AI data and model sharing. In 2022, the GPAI IP Project Advisory Group focused on identifying the challenges confronting these efforts as well as possible solutions. This work was informed by research, interviews and consultations with various stakeholders, and consultation among members of the GPAI Multi-stakeholder Expert Group.

The [Preliminary Report on Data and AI Model Licensing](#) (the “Preliminary Report”) summarizes the findings of the GPAI IP Project Advisory Group’s work on this topic in 2022. This report explains several of the legal, technical, and business challenges confronting the licensing of AI data and models. It also highlights many steps for addressing the challenges. These steps include developing an inclusive ecosystem where a broad range of diverse stakeholders can have access to relevant information and work collaboratively to advance standardized contract terms.

Building upon the Preliminary Report, the GPAI IP Project Advisory Group is working to

support the development of an informed and inclusive ecosystem that can advance efforts to develop standard contract terms for voluntary AI data and model sharing. The Preliminary Report provides a foundation for collaboration and further exploration of solutions to the challenges.

To advance this goal, the GPAI IP Project Advisory Group will host two workshops in 2023 focusing on developing standardized license terms for AI data and models. The first workshop will be held in Munich on the 27 and 28 April 2023. The second workshop is planned to take place in Washington, DC in the summer of 2023.

These workshops will serve as a platform for convening a broad range of stakeholders for in-depth, multi-perspective discussions. The overarching objective of the workshops is to increase the understanding of how to overcome some of the challenges identified in the Preliminary Report and to provide information to advance the drafting of licensing terms.

The workshops will be conducted pursuant to the Chatham House Rule to encourage candid and robust dialogue. To ensure that insights from these dialogues are shared broadly, a synthesis of the workshops' outcomes will be made publicly available. This report will be published in 2023, following the workshops.

Workshop participants will include representatives from a variety of stakeholder groups, including policymakers, legal scholars, industry representatives, data scientists, engineers, representatives of non-profit organizations and other experts.

The workshop is by invitation only, and the invitation is personal and non-transferable.

Annex 2: Workshop program

Location: Max Planck Institute for Innovation and Competition (Marshallplatz 1, 80539, Munich)

27 April 2023	
13:30 – 14:00	Registration
14:00 – 17:00 (including a coffeebreak)	Session 1: The current status of the standardization of licenses for AI data and models: Challenges and prospects
	Dinner (Altes Hackerhaus, Sendlinger Str. 14, München, www.hackerhaus.de)
28 April 2023	
9:30 – 13:00 (including a coffee break)	Session 2: The role of potential IP and trade secrets protection in the context of licensing AI data and models The existence of IP protection, especially copyright and neighboring rights protection, or trade secrets protection may be uncertain under most jurisdictions. Potential existence of third- party rights and trade secrets may create obstacles to AI data and model licensing. How can contract terms take account of those risks?
13:00 – 14:00	Lunch
14:00 – 16:45 (including a coffeebreak)	Session 3: AI liability as a potential obstacle to licensing? What impact will potential product and general tort liability have on the use of AI systems under upcoming legislation such as in the EU for AI data and model licensing? How can contract terms take account of such risks of liability?
16:45 – 17:00	Workshop results and an outlook

Annex 3: Discussion Outline

Session 1: The current status of the standardization of licenses for AI data¹⁹ and models: challenges and prospects

Motivation: Data transactions and the development of AI systems are rapidly developing areas of business, institutional practices, and policymaking. As described in the [2022 Preliminary Report on Data and AI Model Licensing](#), several initiatives are underway worldwide to develop standard contractual terms that would address the unique aspects of AI data and models. The first session of this workshop will delve deeper to understand the current state of developing standardized licensing terms, and the progress made since the 2022 Preliminary Report. This would establish the foundation for the subsequent discussion that will focus on the specific substantive legal issues and how to address them in contractual provisions.

Objectives:

- To cast a broad perspective as regards the state of play in the standardization of licenses for AI data and models, including the work being done by various organizations to advance these efforts;
- To map out the key legal issues and business/market factors that characterize the specifics of such licenses;
- To identify key technical, legal, institutional and business challenges and uncertainties that stand in the way of standardizing licenses for AI data and models;
- To identify strategies and pathways for addressing such challenges and uncertainties in connection with the development of licensing terms for AI data and models.

Discussion points:

- Clarifying terminology/key definitions in the context of licenses for AI data and models. The definitions provided by the 2022 GPAI report could serve as a starting point; the glossary will be further clarified and expanded upon by the participating experts.
- Developments ‘on the ground’: What business models and practices are emerging around transactions for AI data and models? How have markets for AI data and models been developing? What are the specifics of licenses for AI data and models? How do licenses for AI models differ from customary software licenses?
- A bird’s-eye view on the standardization efforts: What progress has been made regarding the standardization of licenses for AI data and models more recently? Have new players and initiatives emerged in this area?
- What technical, legal, institutional, and business challenges and uncertainties persist that obstruct the standardization of licenses for AI data and models?
- What are the potential directions, paths and approaches to address such challenges?
- What notable policy trends and legislative developments are likely to impact licenses and transactions for AI data and models, particularly in a cross-border context?

¹⁹ For the purposes of the discussion at the workshop, ‘AI data’ refers to data used for the development of AI models and applications, including training, testing, and validation datasets.

- How can the voluntary licensing of AI data and models be facilitated?

Session 2: The role of potential IP and trade secrets protection in the context of licensing AI data and models

Motivation: The development of AI systems and applications involves many inputs, some of which might be protected under IP law. However, uncertainty often persists regarding the applicability of IP protection to AI data and models and the scope of such protection. Besides, identifying the relevant right holders and securing the relevant permissions for the development of AI systems might be challenging. This session will explore the aspects of managing IP rights in the context of licensing AI data and models from both the licensor's and the licensee's perspectives.

Furthermore, the discussion will examine the role of trade secrets protection in the context of licensing AI data and models. In contrast to exclusive IP rights, which have received significant attention in discussions on data-sharing and AI, the role and implications of trade secrets protection for AI innovation remain largely unexplored. However, trade secrets protection is potentially applicable to nearly all elements of AI development, including data and AI models. In this view, it is crucial to examine how trade secrets protection plays out in contracts on the sharing of AI training data and models.

Objectives:

- To identify uncertainties and challenges related to IP protection, including legal uncertainty about the existence of such protection, faced by both licensees and licensors when licensing AI data and models, and to consider how they can be addressed in model agreements;
- To deepen the understanding of the implications of potential trade secrets protection for the contractual allocation of access and use rights in AI data and models, particularly in the context of cross-border transactions;
- To examine the legitimate interests of trade secrets holders and potential licensees, and to identify how model contractual terms could be crafted to balance these interests.

Discussion points:

With regard to IP rights:

- What uncertainties and challenges related to IP protection do licensees and licensors encounter in the context of licensing AI data and models?
- What are the viable approaches to solving such challenges? How could they be addressed through standardizing contract clauses?
- What role do IP exceptions and limitations play in situations where IP-protected subject matter is used for the purposes of developing AI systems and applications? Can such instruments of access be deemed sufficient?
- Which 'best practices' of IP management are applicable in the context of licensing AI data and models?

With regard to trade secrets protection:

- What makes AI data and models a unique subject matter from a perspective of trade secrets protection?
- Are there examples of jurisdictional variations in the applicability of trade secrets protection to AI data and models and the scope of protection? How should these variations be addressed in cross-border licenses for data and AI models? How can model contract terms address these variations?
- In situations where there is uncertainty about the applicability of trade secrets protection to AI data and models, how can the risk of misusing such uncertainty be mitigated in data and model licenses? What could be a model clause for that purpose?
- What are the specific characteristics of licenses for AI data and models in comparison to non-disclosure agreements?
- How should the scope of legitimate interests of the trade secrets holder and data/model user be defined? Which principles and instruments of contract law can help balance such interests? How could the model contractual terms be devised to provide for a balanced solution?

Session 3: AI liability as a potential obstacle to licensing?

Motivation: During interviews carried out by the IP Committee of the I&C Working Group of GPAI in 2022, participants stressed that liability issues in the context of AI innovation pose significant uncertainty for licensing AI data and models. As the regulatory landscape for AI continues to evolve, the introduction of technical standards and AI-specific compliance and liability rules pose considerable uncertainty in the realm of contractual arrangements. The upcoming EU AI Act²⁰ represents an unprecedented comprehensive regulation that pursues ambitious goals and sets out far-reaching compliance obligations, including with regard to the quality of data used for the development of AI systems. In addition, a new AI-specific liability regime is being established in the EU that aims to address the potential risks and harms caused by or mediated through AI applications.²¹

The enactment of such targeted legislative instruments raises the question of their implications for licensing practices for AI data and models, in particular, whether contracting parties may be able to draft licenses as 'as is' agreements that exclude most or all warranties and liability. Given that these proposals are currently undergoing advanced stages of legislative review and are expected to retain their core provisions, it is essential to consider these developments and their implications for standardized licenses for data and AI models, particularly in a cross-border context.

²⁰ Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (21 April 2021) COM (2021) 206 final, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>.

²¹ Proposal for a Directive of the European Parliament and of the Council on liability for defective products (28 September 2022) COM(2022) 495, <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52022PC0495&qid=1676546081706&from=EN> (adapting product liability rules to new technologies, in particular, AI); Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) (28 September 2022) COM(2022) 496, [https://ec.europa.eu/transparency/documents-register/detail?ref=COM\(2022\)496&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2022)496&lang=en).

Objectives:

- To gain a deeper understanding of the implications of the upcoming EU AI Act and AI-specific liability regime for the contractual allocation of access and use rights in data and AI models;
- To identify key areas of legal uncertainty and explore potential solutions, particularly in the context of transactions involving AI data and models;
- To investigate the potential and scope of standardizing liability clauses in licenses for AI data and models;
- To consider the need for policy recommendations, including for the European legislature, to address the identified challenges and reduce legal and business/market uncertainties.

Discussion points:

- What provisions of the upcoming EU AI Act are likely to have a significant impact on the contractual allocation of access and use rights in AI data and models? What challenges and uncertainties do these provisions pose for licensing practices? How can/should such challenges and uncertainties be addressed in contracts?
- Will the EU AI Act, particularly the obligations regarding data quality and robustness of the training process, make ‘as is’ agreements for data no longer feasible? In which business models and contractual scenarios can ‘as is’ provisions for warranties and liability limitations regarding AI data or AI models be acceptable and lawful?
- Will the EU AI Act, the new AI Liability Directive, and the revised Product Liability Directive require the inclusion of new clauses in the data- or model-sharing agreements? What are the implications of these upcoming acts for cross-border transactions for AI data and models?
- Contracts for AI data and models are often concluded in situations where data is used as input for research and development (R&D), and there is a potential for harm or damage to be caused or mediated by the R&D output. Which considerations should be taken into account when drafting liability clauses in such scenarios?
- How will the statutory and contractual allocation of liability interact in licenses for AI data and models? What are the implications of the EU’s new liability framework for the allocation of liability for harm caused to third parties by products or services that were developed using transacted data or incorporating transacted AI models? How much leeway would parties have to contractually define the scope of liability, liability addressees, types of harm, indemnification, and the applicable substantive law as regards liability?
- To what extent is it feasible and viable to standardize contract terms for defining and allocating liability for harm caused to third parties in licenses for AI data and models?

APPENDIX 2

Exploring Pathways to the Standardization of Licenses for Data and Machine Learning Models

A workshop co-organized by the Global Partnership on Artificial Intelligence, the Max Planck Institute for Innovation and Competition (Munich), and Duke University

Annex 1: Workshop Program

When: 20th and 21st of June 2023.

Where: Duke University in Washington, DC (1201 Pennsylvania Avenue, NW, Suite 500)

The workshop topics address issues raised by generative AI and other AI applications. Annex 1 includes more detailed descriptions of each workshop discussion and lists materials that participants should review in advance of the workshop. Annex 2 broadly describes the work of the GPAI AI Advisory Committee.

Day 1	
12: 30 – 13:00	Registration (refreshments and light snacks provided)
13:00 – 13:05	Welcome Remarks.
13:05-13:30	Presentation and Discussion by Russell Hanser, Associate Director National Telecommunications and Information Administration, US Commerce Department.
13:30-14:00	Participant Introductions.
14:00 – 17:30 (Including a coffee break)	1 st Session: Using Contracts to Address AI Data Scraping, Text and Data Mining, and the Use of Publicly Accessible Computer Code for AI Applications. The scraping and use of publicly accessible data and code has skyrocketed with the rise of large language models and generative AI, but the corresponding legal landscape remains very uncertain. How might standardized contract terms provide more certainty for these practices? This session will include a short presentation by David Nimmer.
18:30	Workshop participants are invited to attend a dinner sponsored by ML Commons at Clyde's, 707 7 th Street, NW
Day 2	
8:00-8:40	Buffet breakfast available
8:40-8:45	Day 2 Welcome Remarks
8:45-9:00	Emmanuel Barcy, Chief Scientific Officer, French Data Health Hub, Virtual Presentation

9:00-10:00	<p>2nd Session: Contractually Allocating Rights to Generative AI Prompts, Trained AI Models and Their Outputs and Addressing Imbalances in Bargaining Power</p> <p>Input data for developing AI, AI models, generative AI prompts and outputs have tremendous value. How might standard contract terms be formulated to facilitate fair and efficient allocation of usage rights in this subject matter? What categories of contractual arrangements are most likely candidates for standardization? Is a menu of standard terms needed to account for the various potential contractual arrangements? To what extent can standard contract terms address imbalances in bargaining power?</p>
10:00-10:30	Leonardo Cervanas-Navas, Director, European Data Protection Supervisor, short remarks followed by Q&A.
10:30-10:50	Coffee Break
10:50 – 12:30	<p>2nd Session (<i>Continuing. See above</i>): Contractually Allocating Rights to Generative AI Prompts, Trained AI Models and Their Outputs and Addressing Imbalances in Bargaining Power</p>
12:30 – 13:30	Workshop participants are invited to attend a lunch on premises sponsored by Microsoft
13:30 – 16:30 (Including a coffee break)	<p>3rd Session: Using Contracts to Address Ethical Concerns and to Comply with Emerging AI Regulations</p> <p>Generative AI and other AI applications increasingly raise ethical concerns and are becoming more regulated. Should contracts help establish guardrails to promote ethical behavior and regulatory compliance, and if so, how? How do available technical tools impact the role of contracts in addressing ethical issues?</p>
16:30-17:00	Wrap-up: Path forward – next steps and how can GPAI be helpful to ongoing efforts?
18:15	Workshop participants are invited to attend a dinner sponsored by Northrop Grumman at the Bombay Club, 815 Connecticut Ave NW, Washington, DC 20006

Annex 2: Discussion Outline and Materials to Review in Advance of Workshop

1st Session: *Using Contracts to Address AI Data Scraping, Text and Data Mining, and the Use of Publicly Accessible Computer Code for AI Applications.*

Motivation: Many AI applications require access to significant amounts of AI data, including for training, testing and validation. Stakeholders are aggregating AI data through a variety of means, including scraping or mining data (data scraping) from third party websites and social media properties. In addition, organizations increasingly are scraping open source and other publicly available computer code in connection with the development of AI models. These practices have skyrocketed, along with the meteoric rise of generative AI.

The legal and policy frameworks for these activities have not kept pace with industry and other stakeholder developments. In response, various organizations have imposed contractual terms to govern scraping and other practices relating to publicly accessible data and code. These contracts and practices have been the subject of litigation, and the legal cases continue to mount.

This session will focus on emerging approaches for developing form agreements or standardized contractual terms for scraping AI data and code, taking into consideration US copyright law, the EU Text and Data Mining Directive, and certain other relevant laws. For example, it will look at existing precedents, such as *LinkedIn v HiQ*, and emerging cases, and consider how standard contract terms might provide more certainty with respect to data scraping and using publicly available code. Drawing upon the Preliminary Report, participants are encouraged to prepare and discuss potential contract terms that would better accommodate the myriad of different types of desired sharing arrangements.

Finally, this session will focus on how standardized TDM and scraping contract terms can potentially advance the important goal of international harmonization, since data and technology do not stop at geographic borders. The EU, the US, and other jurisdictions have different intellectual property laws. In the US, one consideration is whether a particular scraping practice constitutes a permitted fair use under US copyright law. In contrast, there is no fair use exception to EU copyright protection. However, the EU Text and Data Mining Directive provides an alternative framework for managing these activities. Consideration will be given to structuring contract clauses that might facilitate cross-border activities, given the differing legal regimes.

Discussion Questions:

1. A potential approach for leveraging contracts is to create an ecosystem of multiple parties who grant reciprocal rights to each other for the scraping of certain data for certain uses. To what extent is this approach feasible?
2. Are there royalty-free terms that people would agree to?
3. With a contractual approach to data scraping, what should be the scope of permitted use? Could this scope include personally identifiable information, and if so, what type? If so, how would the data use be limited?
4. What other restrictions should apply to the use of scraped data or code? Should it comprise only a de minimis portion of the training data? Should there be requirements that the AI

outputs are lawful and are not the same as or substantially or confusingly similar to the scraped data or code or have the same or a similar likeness? Should sublicenses generally be prohibited?

5. How might contracting terms for data scraping differ from code scraping? Do the considerations for code scraping differ?
6. How can common data scraping terms advance international harmonization? What are the potential hurdles?
7. How might a royalty sharing structure be created for situations where the use is not de minimis or the outputs are similar to the inputs?

Pre-reads for Workshop Session 1:

1. [intellectual-property-expert-preliminary-report-on-data-and-AI-model-licensing.pdf \(gpai.ai\)](#) (GPAI IP Committee 2022 Report)
2. [Should CC-Licensed Content be Used to Train AI? It Depends. - Creative Commons](#)
3. [Generative AI Debate Braces for Post-Warhol Fair Use Impact \(bloomberglaw.com\)](#)
4. [Japan Goes All In: Copyright Doesn't Apply To AI Training \(technomancers.ai\)CC Community Input: Better Sharing for Generative AI - Creative Commons](#) (optional)

2nd session: Contractually Allocating Rights to Generative AI Prompts, Trained AI Models and Their Outputs and Addressing Imbalances in Bargaining Power.

Motivation: In agreements on data and AI model sharing, it is common that the parties define the kinds and scope of use permitted to the recipient. Especially where the subject-matter may not be protected by IP or trade secrets law, the provider may impose any use restriction on the recipient. Similar issues can arise in the context of generative AI if the provider of generative AI applications claims extensive rights of use as regards the data included in the prompts and restricts the rights of use as regards the output of generative AI.

This session will therefore focus on the contractual allocation of usage rights in AI input data, models, generative prompts, and output (collectively, 'AI-related subject matter'). The overall goal is to stimulate and exchange ideas regarding how standard contract terms might be formulated to facilitate the allocation of usage rights in AI-related subject matter in particular by addressing legal uncertainties and remedying imbalances in bargaining power. As regards the latter, it should be noted that the EU legislature is currently about to adopt rules to control unilaterally imposed B2B contract terms in data sharing agreements including contract terms containing data use restrictions (see attachment on the current proposals, also referenced in the pre-read materials).

Discussion Questions:

1. AI data, models, prompts, and output are shared in various contexts (B2B, B2C, C2B) and under different conditions and modalities. Which **typical scenarios** are the most likely candidates for the standardization of contracts?
2. How can standard definitions be developed to support the creation of standard contract terms?

3. What are the specific **legal issues** and **uncertainties** related to the allocation of access and usage rights in AI data, prompts, models, and output that can be and need to be addressed in standardized contracts?
4. AI-related subject matter can be utilized in various ways. Is there a need for a non-exhaustive list of specific usage rights and their common/standardized definitions? What rights should such a list include? What are the main dimensions of usage rights in AI-related subject matter that need to be defined in contracts?
5. Contracts often allocate specific usage rights to AI-related subject matter. These allocations are often characterized by inequalities in bargaining power among the prospective parties. What are the main **sources of inequalities** in bargaining power (e.g., data dependence, uniqueness/un-substitutability of data(sets), use of indispensable applications such as certain large language models)? Do such differences require tailored solutions? To what extent can voluntary model terms remedy such inequalities? When should mandatory contract law step in?
6. Contractual arrangements for sharing AI-related subject matter might be increasingly subject to (evolving) **mandatory contract law** in some jurisdictions. Among the novelties in the EU is the introduction of fairness control of data-sharing agreements under the proposed EU Data Act. To be discussed: Drawing on the EU Data Act as an example, to what extent can mandatory rules on fairness be desirable in AI data-sharing scenarios (and more broadly, transactions over AI-related subject matter)? To what extent should fairness rules inform standard contract clauses for AI-related subject matter?
7. Legal uncertainty often arises due to potential third-party rights in AI-related subject matter. Which strategies could be employed by contracting parties to navigate uncertainty about the applicability of existing laws the existence of erga omnes third-party rights in AI-related subject matter?

Pre-reads for Workshop, Session 2:

1. [When AI generates work, standard contractual terms can help generate value and clarity - OECD.AI](#)
2. [Excerpts from draft EU Data Act in attached pdf.](#)

3rd Session: Using Contracts to Address Ethical Concerns Raised by Generative AI and other Applications and Related Regulatory Compliance and Standard Contract Definitions.

Motivation: Generative AI and other AI applications have sparked significant ethical concerns. As discussed in the Preliminary Report, there are efforts underway to develop contractual terms to prohibit unethical or illegal uses of AI applications. Additionally, in anticipation of the enactment of the EU AI Act, parties are already considering how to impose contractual terms in AI data and model sharing agreements to comply with the expected regulatory requirements, such as post-market surveillance requirements for high-risk AI, output. This session will discuss the issues broadly and focus on approaches set forth in the contracts included in the pre-read materials. It also will consider mechanisms for effectively enforcing contractual terms.

Discussion Questions:

1. Do ethical limitations or use restrictions belong in license agreements?
2. To what degree can technical tools be a substitute to ethical licensing terms? Should tooling be addressed in a license agreement?
3. Should license agreements include disclosures about the quality and/or accuracy of data? How can this be done?

4. In defining output limits and purposes, what should be the defining scope?
5. In comparing the different approaches in the RAIL and Linux licenses, what are the strengths and weaknesses of each license?
6. What other licenses exist or are being developed that are relevant to this effort?

Pre-reads for Workshop, Session 3:

1. [Responsible AI licenses: a practical tool for implementing the OECD Principles for Trustworthy AI - OECD.AI](#)
2. [BigCode Open RAIL-M v1 License Agreement – Hugging Face](#)
3. [AI Pubs Research-Use RAIL-M License – Project RAIL](#)
4. [Community Data License Agreement – Permissive – Version 2.0 – Linux Foundation](#)
5. [About CC Licenses - Creative Commons](#)
6. [Usage Policies - OpenAI](#)
7. [Anthropic Legal Center](#)
8. [Generative AI Additional Terms of Service \(google.com\)](#)
9. [MLCommons](#) (see website)
10. [Open Data Collaboration and Sharing | Microsoft CSR](#) (see legal frameworks)
11. [Casual Conversations Dataset Licensing Agreement – Meta](#)

Annex 3: Background Note on GPAI AI Advisory Committee's Work

In recent years, it has become widely acknowledged that access to data is crucial for artificial intelligence (AI) innovation and that voluntary data sharing can help address this need. AI depends on appropriate data for many purposes, including training models, testing, and validation (collectively, 'AI data'). Interest also has increased in expanding tools for the voluntary sharing of AI models, whether on an open-source basis or pursuant to other terms. Despite these pressing needs, society has not yet achieved optimal levels of responsible and efficient AI data and model sharing due to several legal, technical, and economic challenges. It is therefore imperative for stakeholders, including companies, governments, academia, and non-profit organizations, to work collaboratively to address these challenges with the goal of facilitating AI data and model sharing.

There is emerging consensus that standardized contract terms can help overcome some of these challenges by reducing transaction costs, fostering more legal certainty, and helping to advance other related goals. Relatively nascent efforts are underway to develop such standardized terms.

The IP Project Advisory Group of the Innovation and Commercialization (I&C) Working Group of the Global Partnership on Artificial Intelligence (GPAI) recognizes the importance of fostering more responsible and efficient AI data and model sharing in order to help unlock the promise of AI. For this reason, the GPAI IP Project Advisory Group launched a project in 2022 to assess and support ongoing initiatives to develop standard contract terms for AI data and model sharing. In 2022, the GPAI IP Project Advisory Group focused on identifying the challenges confronting these efforts as well as possible solutions. This work was informed by research, interviews and consultations with various stakeholders, and consultation among members of the GPAI Multi-stakeholder Expert Group.

The [Preliminary Report on Data and AI Model Licensing](#) (the "Preliminary Report") summarizes the findings of the GPAI IP Project Advisory Group's work on this topic in 2022. This report explains several of the legal, technical, and business challenges confronting the licensing of AI data and models. It also highlights many steps for addressing the challenges. These steps include developing an inclusive ecosystem where a broad range of diverse stakeholders can have access to relevant information and work collaboratively to advance standardized contract terms.

Building upon the Preliminary Report, the GPAI IP Project Advisory Group is working to support the development of an informed and inclusive ecosystem that can advance efforts to develop standard contract terms for voluntary AI data and model sharing. The Preliminary Report informs this ecosystem and provides a foundation for collaboration and further exploration of solutions to the challenges.

To advance this goal, the GPAI IP Project Advisory Group will host two workshops in 2023 focusing on developing standardized license terms for AI data and models. The first workshop will be held in Munich on the 27th and 28th of April 2023. The second workshop will take place in Washington, DC on the 20th and 21st of June 2023. The second workshop will build upon the first workshop, with a particular focus on the standardization of licenses and related challenges involving large language models or generative AI. These workshops will serve as a platform for convening a broad range of stakeholders for in-depth, multi-perspective discussions. The overarching objective of the workshops is to increase the understanding of how to overcome some of the challenges identified in the Preliminary Report and to provide information to advance the drafting of licensing terms.

The workshops will be conducted pursuant to the Chatham House Rules to encourage candid and robust dialogue. To ensure that insights from these dialogues are shared broadly, a synthesis of the workshops' outcomes will be made publicly available. This report will be published in 2023, following

the workshops.

Workshop participants will include representatives from a variety of stakeholder groups, including policymakers, legal scholars, industry representatives, data scientists, engineers, representatives of non-profit organizations and other experts.

The workshop is by invitation only, and the invitation is personal and non-transferable.

Appendix 3: When AI generates work, standard contractual terms can help generate value and clarity



[Lee Tiedrich](#)

[Distinguished Faculty Fellow in Ethical Technology, Duke University
Science & Society](#)



Works generated by Artificial Intelligence (AI) can have tremendous economic value. Back in 2018, [Christie's sold an AI-generated painting](#) for \$432,500, far exceeding the sales price of authentic Warhol and Lichtenstein works displayed in its gallery. Since then, [AI-generated art has blossomed](#), sparking debates about whether Intellectual Property (IP) laws should protect AI-generated works and related developments and, if so, how such rights should vest.

The questions raised by AI-generated works and related developments [extend](#) well beyond art and impact many facets of society. [Journalism](#), [advertising](#), [music](#), [biotech](#), and [other](#) sectors face similar challenges. For instance, AI has improved drug discovery, stimulating further AI adoption and investment in this industry. AI produces significant amounts of computer code. It can also help design computer chips, industrial parts, and materials and generate synthetic data, which is critical for certain Privacy Enhancing Technologies (PETS). These AI-generated outputs can have significant economic value, as can the underlying AI models, the individual data sets used to train, test and validate such models, the aggregate training, testing and validation data sets, and the prompts used to query such models.

While the adoption and valuation of large language models (LLMs) and other AI developments continues to soar, IP laws do not clearly answer critical questions, such as who owns and has the right to use such works. The questions become even more complicated when multiple organizations contribute significant expertise and resources to create “co-generated AI works.” This uncertainty can deter responsible innovation because it can make the return on investment much less predictable. It also increases [litigation risks](#).

Fortunately, contracts can help clarify the ownership and permitted uses of AI works and reduce risks, even when underlying [IP laws remain uncertain](#). Contracts already fill this gap in many circumstances. However, to optimize their benefits, standard contract terms are needed. As explained in a recent [report](#) published by the Global Partnership on AI (GPAI), standard contract terms can reduce transaction costs, provide greater certainty for more organizations, and address an array of other issues. Developing common contractual definitions would advance this work.

IP challenges with co-generated AI works

Policy makers worldwide have encouraged responsible AI innovation, and many organizations are combining resources to create co-generated AI works. While expanding AI development, this raises novel and complex IP issues. For instance, an organization might produce an AI model to help design automotive brakes. The model may include the organization's proprietary code as well as third-party code. To train the AI model, the organization might rely upon its own data, proprietary third-party data, and publicly available data, at least some of which are [scraped from third-party platforms](#). It may then use different third-party data to test and validate the trained model. The organization may invest significant resources in performing data hygiene, developing APIs, and otherwise aggregating and preparing the data sets for use with its AI models.

Once trained and validated, the AI model may be licensed to the organization's customers, who use prompts to query the model. The customers may experiment with their own prompts as well as third-party prompts. "[Prompt engineering](#)" has emerged as a new field. Queries based upon these prompts may then result in patentable brake components.

This hypothetical highlights the challenges of identifying the inventors and owners of co-generated AI works. To what degree should the organization, its code and data suppliers, its customers, and their respective third-party prompt engineers [have a stake](#) in the resulting patentable brake components? What interests should they have in the trained AI model or the aggregate data sets used to train, test, and validate the AI model? What happens if two or more entities produce the same prompts or the same patentable results? How might trade secrets or copyright laws protect AI prompts? How might trademarks and trade secrets be leveraged to enhance protection? AI value chain participants will want answers to these questions as they decide how best to engage in AI innovation.

The evolving IP legal landscape

Unfortunately, today's IP laws are not well-equipped to provide definitive answers to these critical questions. As a threshold matter, countries differ on whether and to what extent AI-generated works qualify for IP protection. This can have important commercial implications, given the value typically associated with IP.

With respect to patents, many jurisdictions currently require human inventors, including the [United States](#), [Australia](#), the [United Kingdom](#), the [European Patent Office](#), [Germany](#), [Brazil](#), and [Israel](#). [South Africa](#), in contrast, recognizes AI patent inventors. Even with this guidance, the patent legal landscape remains unsettled. An AI inventorship case is pending before the [UK Supreme Court](#). In the US, the Patent and Trademark Office launched a [series of public engagements](#) on AI and inventorship.

The legal uncertainty also extends across the copyright landscape. Several jurisdictions, including the [United States](#), Korea, [Australia](#), and the [European Union](#), require human authorship of copyrighted works. However, the level of required human involvement for copyright protection remains somewhat unsettled. For instance, the [US Copyright Office](#) recently denied copyright

protection for AI-generated images in the comic book [Zarya of the Dawn](#), but confirmed that the human author has copyright protection for the text as well as the “selection, coordination, and arrangement of the work’s written and visual elements.”

As part of its recently launched [AI Initiative](#), the US Copyright Office [issued guidance](#) on the level of human authorship required to obtain copyright protection. Among other things, this guidance indicates that when an AI model receives a directional prompt, the resulting work will not have copyright protection if it lacks sufficient human creative control. The US Copyright Office plans to seek further public input soon. Meanwhile, other jurisdictions, such as [Canada and India](#), have entertained copyright co-authorship for works developed jointly by humans and AI.

Contracts can increase certainty

Parties entering AI collaborations typically want clarity upfront on their ability to own and control the fruits of their creative efforts. Fortunately, contracts can reduce some uncertainty presented by IP laws, which is important given the dramatic rise of AI-generated works. Specifically, contracts can help organization’s structure collaborations to include sufficient human involvement, reducing the risk of forgoing IP protection. They can also include confidentiality and other provisions to help secure trade secret protection, which may be preferred, particularly when patent protection seems too improbable or expensive. Additionally, contracts can override some default rules and ambiguities under current IP laws with clearer and mutually agreeable terms. This practice is commonplace in many technology transactions, including some AI terms of use agreements. For instance, Open AI’s [Terms of Use](#) state:

“You may provide input to the Services (“Input”), and receive output generated and returned by the Services based on the Input (“Output”). As between the parties and to the extent permitted by applicable law, you own all Input. Subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title, and interest in and to Output.”

Creating standard AI contract terms

Developing standard contract terms can extend greater certainty and other benefits to more organizations, including Small and Medium-Size Enterprises (SMEs). Importantly, it can help address bargaining power inequities by offering organizations a menu of diverse options for contractually allocating rights, responsibilities, and risks. Additionally, it should lower transaction costs by reducing the need to develop novel and often complicated contractual terms that typically lead to lengthy negotiations.

Standard contract terms can also potentially increase open data sharing, like Creative Commons and Open-Source standard licenses have facilitated royalty-free content and software sharing. These benefits should enhance competition by decreasing barriers to entry, another important societal goal. They can also advance policy efforts to foster greater data sharing, such as the [US Federal Data Strategy](#), the [EU Data Governance Act](#), and the [New Zealand Net-Zero Data Public Utility](#).

Definitions

As discussed in the GPAI Report, to advance standardization, attention must increase on developing common contractual definitions. The definitions should take into consideration various elements of the “AI co-generation” value chain, including the following.

- Original input data. A contractual term may be needed to refer to data in the form first made available for the AI collaboration. Parties may want to have different terms to distinguish

between original data provided for AI model training and original input data provided for other purposes.

- Processed data set. One or more terms may be needed to define the data set developed using original input data. As explained in the GPAI Report, this could potentially encompass: i) a cleansed data set developed by performing data hygiene on original input data and ii) any data compilation, database, insights, or metadata developed using original input data, either in its original form or after undergoing data hygiene.
- Untrained model. Parties may want to have a definition for an AI model that has not been trained.
- Trained model. Parties may want to have a definition for a trained AI model and/or specific parameters, such as weights.
- Trained model outputs. This definition could refer to the outputs of the trained AI model, such as the patentable automotive brakes discussed above.
- Prompts. This definition could refer to prompts created to query the trained AI model.

Multi-stakeholder collaboration

Developing standard contractual definitions and other terms requires multi-disciplinary collaboration. This will help ensure that the resulting legal terms align with the desired technical, business, and ethical approaches. To foster fairness, the drafting process should also integrate the viewpoints of different stakeholders, including SMEs and historically under-represented people. Where relevant, it also should draw upon existing legal practices and emerging policy developments, such as the [OECD Framework for the Classification of AI Systems](#) and potentially government procurement rules. This inclusive process should lead to better results and quicker adoption.

Stakeholders should get involved

Given the rapid pace of AI developments, there is no time to waste in developing standard contract terms that can lower barriers to entry and provide more certainty for AI innovators seeking to unlock the benefits of responsible data and AI. However, much work remains to be done, including with respect to standardizing definitions. Organizations that are interested and have relevant experience or perspectives should contribute to the process. The GPAI Report provides important information on how to engage. This work also has the potential to help inform policy makers about societal needs as the IP legal landscape continues to evolve.

Thanks to [Alban Avdulla](#), a recent LLM graduate of Duke Law School, for his research assistance.