# How Gov't AI Protections May Affect Contractors' Data Rights

By **Tyler Evans and Caitlin Conroy** (August 15, 2024)

Tyler Evans

Caitlin Conroy

On July 8, the U.S. Senate released a proposed version of the National Defense Authorization Act for fiscal year 2025 that includes a number of provisions indicating the U.S. government will continue to increase its focus on maintaining data rights when contracting for artificial intelligence.

In particular, the draft legislation would require that federal budgets include the cost of obtaining datasets to continually train and improve government AI, which could ultimately force contractors to roll the cost of datasets into the price of AI systems, much like the now-ubiquitous approach of requiring that traditional technical data deliverables be identified on proposals and contracts as "not separately priced."[1]

Under the legislation, contractors would also be prohibited from independently developing AI using data provided by the U.S. Department of Defense.[2]

In light of this development, contractors should be aware of how data rights apply when using or providing AI in the performance of a government contract. The current government data rights framework was designed for hardware specifications and early forms of software, and there are unique considerations that arise when it is applied to AI.

**Databases are distinct from software.**

When it comes to data rights, for example, computer databases and software are subject to different frameworks. Databases are not considered part of the software with which they are used and instead can be considered technical data, similar to engineering schematics or manufacturing processes.[3]

As a result, unlike software, the government can obtain the right to use a database within the government for any purpose if it is merely used to perform a contract with a civilian agency.[4]

Databases must be delivered to the government for this right to be usable under contracts with the DOD, but delivery can occur in this context through a variety of unexpected avenues, such as through unrestricted government audit rights involving quality or vulnerabilities.[5]

The distinction between a database and software matters for AI because datasets qualify as a form of database, and are critical to AI development and, at times, operation.[6]

High-quality, curated datasets can be used to train high-quality AI. For example, under a large language model, curated datasets increase the quality of responses to user queries and reduce hallucinations.

In addition, datasets that are continually updated through an online AI system can be even

more valuable because they reflect updated content and operational input from users, rather than merely static data that a single developer thinks will adequately train an AI model.

Rights in datasets can, therefore, enable the government to develop its own quality AI models even if it does not have access to the source code of other models trained on the same dataset. Unlike traditional software, access to datasets can go a long way to establish similar functionalities in competing AI. For example, a civilian agency could potentially leverage a database of claims and resolutions to continually improve AI in the management of an entitlement program by training its own AI model with similar functionalities.

Agencies could also leverage a wide range of other datasets to achieve the same result, such as those consisting of facial recognition, flight and off-road environments, communications, cyber threats, and other national security data.

The Defense Innovation Board has recognized this fact, and has even encouraged the DOD to view data as a product and start compiling catalogs of data for use in additional government projects.[7] The Office of Management and Budget also recently issued governmentwide guidance making similar recommendations.[8]

Although contractors can often avoid the application of default government data rights by providing items that qualify as "commercial" under applicable regulations, that datasets are not considered software limits the usefulness of this approach.

For example, civilian agencies can be required to accept commercial licenses for databases and software, but default rules regarding the government's minimum rights in databases would generally continue to apply and likely take precedence.[9]

The DOD offers additional protections for databases, but would still be in a position to push for greater rights if there are any modifications to a commercial dataset that occur in performance.[10]

Moreover, the DOD has historically adopted an extremely broad view of the types of data that are necessary for "operation, maintenance, installation, or training" under Defense Federal Acquisition Regulation Supplement 252.227-7015, which are subject to unlimited government rights.[11] The DOD could try to extend this view to override commercial licenses for certain datasets even though a similar approach would not be possible for software.

As a result, those using AI in performance of a government contract should carefully consider whether they need to take steps to protect valuable datasets associated with the AI, such as by only offering AI models in offline — i.e., static — modes, or by making clear that any updates to datasets will only take place outside the scope of a contract.

In addition, contractors should consider whether there are any deliverable requirements that could cover datasets, such as data-ordering clauses, broad contract data requirements lists or other descriptions of deliverables. Contractors may be able to modify these elements to prevent the government from accessing datasets even if it theoretically has the right to use them.

DOD contractors in particular should also consider whether they can assert that relevant datasets do not consist of technical or scientific information, thereby preventing the government from obtaining any rights.

Unlike civilian agencies, the DOD is subject to a data rights regime that typically only grants rights in technical data or software, and leaves open the possibility that a computer database may consist of information that does not fall into either category.[12] Accordingly, something like the dataset of claims and resolutions described above may fall outside the scope of the DOD's rights.

**Code needs to be traceable to limit government rights.**

Separately, when developing or modifying AI under a government contract, contractors should be aware that the context in which individual lines of code are developed often needs to be traceable to limit the government's rights. The government typically obtains unlimited rights in software delivered by a contractor unless the contractor can demonstrate that the software was developed at private expense.[13]

Under DOD contracts, delivery is not necessarily required for the government to obtain these rights.[14] Accordingly, being able to identify when code is developed or modified under a contract can be critical to protecting an AI model.

For DOD contracts in particular, DFARS Section 227.7203-4(b) provides that determining whether software is developed at private or government expense occurs at the "lowest practicable segregable" level, such as by individual subroutine.[15]

If a subroutine is first generated in the performance of a contract and shown to be reasonably expected to perform its intended purpose, it will not be deemed to be developed at private expense, and the government will often obtain unlimited rights.[16]

Similarly, if a subroutine is merely modified in performance of a contract and is subject to the same expectations, the government will often obtain so-called government purpose rights, which authorize use for any government purpose, including with competing contractors.[17]

Many contractors may not be concerned about the DOD obtaining rights in a few subroutines, but tracing the funding of any code that is developed through machine learning can be extremely difficult, if not impossible, without built-in, explainable AI processes.

Specifically, when developing or modifying black-box AI under a government contract, the extent to which new subroutines are created, or existing subroutines are modified, may not be clear.

For example, training an artificial neural network on a new dataset under a contract may result in changes to a relatively low number of neural nodes. However, a contractor may not have adequate documentation to demonstrate this fact.

Contractors generally have the burden of demonstrating that software was developed at private expense, and it may not be possible to identify segregable components of black box AI when the government maintains that the AI was modified in performance of a contract.[18]

In addition, contractors may not be able to satisfy standard obligations in subsequent DOD contracts to identify components of code in which the government already has rights — and to avoid charging the government again for these rights — which raises concerns about misrepresentations and false claims if code is not sufficiently traceable.[19]

This problem is exacerbated under civilian agency contracts because applicable regulations do not expressly adopt the same lowest-segregable framework that applies to the DOD. Instead, civilian agency regulations arguably contemplate that the government can obtain unlimited rights in delivered software if there is even the slightest modification to code in performance, such as through updated parameters resulting from a machine-learning process.[20]

Thus, even if a contractor can prove how code was funded, any use of nonstatic AI in a civilian agency contract risks granting the government unlimited rights to the extent that there is a provision for the AI's delivery.

Fortunately, unlike for databases, default rules on the tracing of funding can be modified for AI models that qualify as commercial computer software, as long as a contract includes a commercial license and omits or limits standard government data rights provisions.[21]

To qualify for this status, software generally needs to have been sold, licensed, or offered to the public or, for civilian agencies, offered as a service that meets a similar market test for similar software.[22]

Importantly, software can continue to retain its commercial status even if it is subject to minor modifications under a government contract.[23] For noncommercial software, DOD contractors can also sometimes negotiate a special license that deviates from the standard requirements.[24]

**Restrictions on using contract data to improve AI will likely increase.**

Contractors looking to use contract data to train AI should also be aware that the government is increasingly restricting this practice. By default, contractors generally have the right to use data that they generate in performance of a contract, and historically have only been subject to obvious limitations on using data received from the government, such as national security or privacy concerns.[25]

However, consistent with the Senate's current version of the 2025 NDAA, agencies are increasingly restricting how data generated or received under a contract can be used in other contexts, particularly if non-U.S. citizens or companies are involved.

For example, civilian agencies can include language limiting a contractor's ability to use AI output and the queries generated in performance of a contract. The DOD also frequently relies on a provision prohibiting disclosure of any information pertaining to a contract or any related program. The provision can also be supplemented by a restriction on a contractor's internal use as well.[26]

The OMB's recent guidance on AI expressly encourages agencies to pursue these restrictions for federal information, which could apply to almost any contract data because this concept covers information created, collected, maintained, disseminated, disclosed, disposed of, or even processed by or for the government.[27]

The Defense Innovation Board is pushing a similar approach, suggesting in a January report, "Building a DOD Data Economy," that the DOD should be permitted to claim an ownership interest in data generated through DOD-funded technologies and secure "expansive rights for future transformations and data ensembles," which would potentially apply long-term restrictions to contract data and improvements to AI using such data.[28]

Moreover, in addition to the above-mentioned restriction on using DOD data to independently develop AI, the Senate's version of the 2025 NDAA includes another provision that goes further to prohibit DOD contractors from trading in datasets that include personally identifiable information about DOD personnel, even if that data is obtained through unrelated commercial channels.[29]

As a result, contractors that plan to improve AI using contract data, or even data ultimately sourced from the government in other contexts, should consider whether their contracts restrict this practice going forward. Absent careful consideration of this issue upfront, it may be difficult, if not impossible, to limit how contract data is used once it is included in a dataset for training AI models.

The government is also likely to take an unkind view of contractors that fail to maintain sufficient documentation of how contract data is used when subject to these restrictions.

---

*Tyler Evans is a partner and Caitlin Conroy is an associate at Steptoe LLP.*



[1] S. 4638, 118th Cong. § 1620 (2024).

[2] S. 4638, 118th Cong. § 810(c).

[3] See FAR 52.227-14(a); DFARS 252.227-7013(a).

[4] See FAR 52.227-14(g)(4), Alt III; FAR 52.227-16(a).

[5] See DFARS 252.227-7013(b)(3); compare FAR 52.227-16(a), with DFARS 252.227-7026. A more detailed discussion of concerns about delivery of AI in the national security context is included in: Steptoe LLP, Artificial Intelligence and the Landscape of US National Security Law 36 (2024), available at https://www.steptoe.com/a/web/27TqRiFJ6ksAo4qdyKDGZ5/steptoe-white-paper-ai-and-national-security-law-july-2024.pdf.

[6] See FAR 2.101; DFARS 252.227-7014(a)(2).

[7] Def. Innovation Bd., Building a DOD Data Economy 16 (2024), available at https://innovation.defense.gov/Portals/63/20240118%20DIB%20Data%20Economy%20Study_Approved-compressed_1.pdf.

[8] Office of Mgmt. and Budget, Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence 12–13 (2024), available at https://www.whitehouse.gov/wp-content/uploads/2024/03/M-24-10-Advancing-Governance-Innovation-and-Risk-Management-for-Agency-Use-of-Artificial-Intelligence.pdf.

[9] See FAR 12.211; 52.227-14(g)(3), Alt II.

[10] See DFARS 227.7103-6(a).

[11] See DFARS 252.227-7015.

[12] See DFARS 252.227-7013(b), 252.227-7014(b).

[13] See FAR 52.227-14(a), (b)(1)(iv); DFARS 252.227-7014(b)(1).

[14] See DFARS 252.227-7014(b)(1).

[15] See DFARS 227.7203-4(b).

[16] See DFARS 252.227-7014(a)(7)–(8), (b)(1)(i).

[17] See DFARS 252.227-7014(a)(7), (a)(10), (b)(2)(i).

[18] See, e.g., FAR 27.404-5; DFARS 227.7203-10(c)(1), 252.227-7037(c).

[19] See DFARS 252.227-7014(b)(5), (j), 252.227-7017.

[20] See FAR 52.227-14(a), (b)(1)(iv); Ervin & Assocs., Inc. v. United States, 59 Fed. Cl. 267 (2004).

[21] See FAR 12.212, 27.405-3; DFARS 227.7202.

[22] See FAR 2.101; DFARS 252.227-7014(a)(1).

[23] See FAR 2.101; DFARS 252.227-7014(a)(1).

[24] See DFARS 252.227-7014(a)(13), (b)(4).

[25] See FAR 52.227-14.

[26] See DFARS 252.204-7000.

[27] See supra note 6 at 25, 28; OMB Circular A-130 §10.

[28] See supra note 5 at 16.

[29] S. 4638, 118th Cong. § 811 (proposing to amend 10 U.S.C. § 4662).