KPMG

HOT TOPIC

Software data costs

Accounting considerations

February 2026

This Hot Topic outlines the accounting for the costs of data used in and by software, including in the training/re-training of AI software.

💡 Introduction

The meteoric rise of AI has put particular focus on data's role in the training (and re-training) of AI models. Data is fundamental to AI software development, continuous improvement and maintenance. Large datasets are obtained/collected (e.g. licensed from unrelated data providers) and used to train underlying AI models, enabling them to recognize patterns, make predictions and perform tasks with a high degree of accuracy. Model retraining (also often referred to as continuous learning) allows an AI model to adapt to new information, improve its performance and stay relevant in dynamic environments. Model retraining generally requires new data, which may be obtained organically by the AI model through its own interactions and feedback or from external sources. Without regular retraining, an AI model risks becoming outdated, less accurate and potentially unreliable in producing outputs or making decisions or predictions.

The use of data is also fundamental to nearly all types of software (AI or otherwise). For example, data is a crucial input to the key outputs (e.g. reports, analytics, visualizations) of many software applications.

In this Hot Topic, we discuss the accounting for data costs incurred to train and retrain AI software models, as well as the costs of data acquired for use *by* an entity's software (e.g. as an input to produce reports or other outputs that display or analyze that data) but not involved in software training or retraining.

📋 Accounting for data costs: the basics

For the most part, data is either internally generated or acquired via licensing agreement. Entities generally acquire *rights to use* data from third parties rather than purchase data outright. US GAAP does not contain *specific* guidance on the accounting for data. The FASB's Emerging Issues Task Force (EITF) discussed the accounting for database content acquired to which others would be sold access in Issue No. 00-20, *Accounting for Costs Incurred to Acquire or Originate Information for Database Content and Other Collections of Information*. However, no consensus was reached, and the FASB has not subsequently addressed the topic.

In the absence of specific guidance, an entity follows the general guidance that applies to intangible assets more broadly. Thereunder, costs to (1) generate data solely for a particular R&D project or (2) license data solely for an R&D project and for which there is no alternative future use are expensed as incurred under Subtopic 730-10 on R&D. Subtopic 350-30 applies to any other internally generated or licensed data. Under Subtopic 350-30, internally generated data costs are generally expensed as incurred, while licensed data is capitalizable, initially measured at its cost to the entity (including any transaction costs), because it arises from identifiable, contractual-legal rights granted by the data licensor. [350-30-25-1, 25-3 – 25-4, 30-1; 730-10-25-2(c)]

Once licensed data is recognized, Section 350-30-35 governs its subsequent accounting. The licensed data is amortized over its 'useful life' to the entity, which is the period over which the data is expected to contribute directly or indirectly to the entity's future cash flows, in a manner that reflects how the entity will consume its benefits. Useful life is not the same concept as economic life or productive life; useful life is an entity-specific determination, which may involve judgment, about the period the asset is expected to contribute to the entity's future cash flows. Paragraph 350-30-35-3 provides a non-exhaustive list of factors that should be considered when estimating an intangible asset's useful life. [350-30-35-1 – 35-3, 35-6]

Licensed data, like any other amortizable intangible asset, is subject to the requirements of Section 360-10-35 on the impairment or disposal of long-lived assets – see KPMG Handbook, Impairment of nonfinancial assets, for guidance. [350-30-35-14]

The basics above notwithstanding, the accounting for licensed data is affected by whether it is used in training AI software. Later sections of this Hot Topic discuss this, based on which US GAAP software development rules apply (see *AI software development*).

Data-related infrastructure costs

An entity will typically incur infrastructure costs to store and use data. These costs may include (not exhaustive): purchased or leased computer hardware (e.g. servers, networking equipment); purchased or leased data center space; and third-party infrastructure (e.g. compute capacity obtained from a cloud service provider) and telecommunications costs, which may or may not include leases not explicitly identified as such (i.e. embedded leases). These costs may be significant for some entities given the extensive data requirements often necessary for AI software development and maintenance.

Such costs are accounted for under Topic 360 (for any purchased facilities or equipment), Topic 842 (for any identified leases), Subtopic 350-40 (for any cloud-based infrastructure service implementation costs) and other US GAAP applicable to service arrangements in general.

Data-related infrastructure costs, such as depreciation of equipment under Topic 360, lease cost recorded under Topic 842, and IaaS service fees, may be direct or indirect costs of an entity's AI software development. Making this distinction is particularly important if the AI software being developed is internal-use software because Subtopic 350-40 treats direct and indirect software development costs differently. See *Internal-use software*.

📋 AI software development

While AI software (in particular, *generative* AI software) is of particular interest (now and likely into the future), the accounting guidance that applies to AI software development is the same guidance that applies to the development of any other software. Therefore, the first step is to determine which US GAAP software development guidance applies.

- **Subtopic 350-40**: Applies to software that is solely for the entity's internal use (internal-use software). Internal-use software can be for true internal-use (e.g. payroll processing, inventory management,

enterprise resource planning) but also includes software to which an entity sells access to customers *solely* on a hosted basis. This latter type of internal-use software includes software an individual or entity may access via an on-device (e.g. smartphone) 'thin-client' application, but that has no (or only very limited) functionality when not connected to cloud-based features or functionalities. Some popular generative and other AI applications qualify as internal-use software because they are only accessed by customers on a hosted basis.

- **Subtopic 985-20**: Applies to costs incurred to develop software to be sold or licensed to third-party customers (external-use software), including software that an entity licenses to a customer as part of a product or providing a service.

- **Other Topics**: Software may be neither internal-use nor external-use; for example, software to be used in R&D is subject entirely to the R&D guidance in Subtopic 730-10, while software an entity develops for others under a contractual arrangement is subject to the contract fulfillment cost guidance in Subtopic 340-40.

Sections 2.3 and 2.4 of KPMG Handbook, Software and website costs, discuss in detail (1) identifying software outside the scope of either Subtopics 350-40 or 985-20 and (2) distinguishing between internal- and external-use software, respectively.

Once an entity appropriately identifies the internal-use, external-use or other nature of the AI software, Subtopic 350-40, Subtopic 985-20 or any other applicable guidance (e.g. Subtopic 730-10 or Subtopic 340-40), is applied to the AI software development without any specific exceptions or different requirements as compared to any other type of software development. KPMG Handbook, Software and website costs, discusses the application of Subtopics 350-40 and 985-20 in detail. Chapter 2 of KPMG Handbook, Research and development, and chapter H of KPMG Handbook, Revenue for software and SaaS, discuss accounting for R&D and contract fulfillment costs, respectively.

While AI software could fall into the 'Other Topics' category described above, we focus the remainder of this Hot Topic on internal- and external-use AI software subject to Subtopics 350-40 and 985-20.

📋 Data not used to train or retrain AI software

The focus of the remainder of this Hot Topic is data licensed to train and/or retrain AI software. The sections that follow discuss this for *External-use software* and *Internal-use software*, respectively.

However, software (AI or otherwise) may access and use data, such as medical or trade data, to produce its outputs (e.g. reports that display or analyze that data), that has no role in training or maintaining (e.g. fixing bugs in) the software's functionality. We believe the US GAAP software development guidance (see *AI software development*) does not apply to the costs of such data. Instead, the general data accounting considerations outlined earlier in this Hot Topic apply (see *Accounting for data: the basics*).
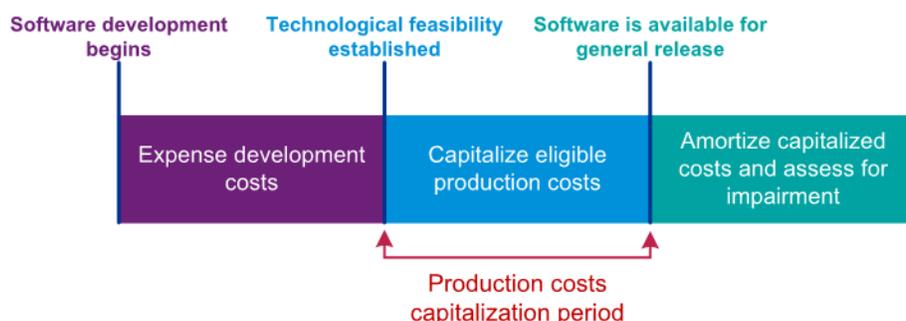
📋 External-use software

Accounting for external-use software development costs

Subtopic 985-20 treats all development costs to establish the technological feasibility of external-use software as R&D costs to be expensed as incurred as required by Subtopic 730-10. [985-20-25-1]

Development (e.g. coding and testing) costs incurred after establishing technological feasibility to produce the 'product master', referred to as 'production costs', are capitalized to the extent recoverable by the net

realizable value (NRV) of the software product until the product is available for general release. [985-20-25-1, 25-3, 25-6]

Software development begins → Technological feasibility established → Software is available for general release

Expense development costs | Capitalize eligible production costs | Amortize capitalized costs and assess for impairment

Production costs capitalization period

Does the acquired data have 'alternative future use'?

The accounting for data acquired for external-use AI software development depends on whether the data has alternative future use. An intangible asset has an alternative future use when (adapted from Question 2.3.80 in KPMG Handbook, Research and development):

- the entity reasonably expects to use it in an alternative manner (e.g. currently identifiable future external-use software development or future maintenance of a different software product) and expects to derive an economic benefit from that alternative use; and

- this alternative use is not contingent on further development of the intangible asset (i.e. the licensed data can be used in the alternative manner in its present condition at its acquisition date).

Data with no alternative future use

Under Subtopic 730-10, the cost of an acquired intangible asset without alternative future use is expensed as incurred. Therefore:

- The costs of licensed data to be used in developing the AI software *before* technological feasibility is established are expensed as incurred, even though such costs may otherwise qualify for capitalization under Subtopic 350-30 (see *Accounting for data costs: the basics*). [730-10-25-2(c)]

- The costs of licensed data used in the further development of AI software *after* technological feasibility is established would, we believe, generally qualify as 'production costs' of the AI software under Subtopic 985-20 and therefore be capitalized. [985-20-25-1, 25-3]

As discussed in section 5.2 and Question 5.3.20 of KPMG Handbook, Software and website costs, for some entities, their software development process may establish technological feasibility only very late in the development cycle, which may lead to very limited, if any, production cost capitalization related to their external-use software.

Data that has an alternative future use

Licensed data with an alternative future use is capitalized and subsequently accounted for under Subtopic 350-30, consistent with the discussion in *Accounting for data costs: the basics*. Given that the licensed data has alternative future use, it may be that the data is being used for multiple purposes concurrently (e.g. to train new AI software and to retrain an existing AI model, to train both an external- and internal-use AI software application, or to train new AI software and serve as content for a non-AI software application). Therefore, amortization may be allocated between the multiple projects/uses, with each allocation accounted for differently (e.g. some allocated amortization capitalized into the cost of a capitalized software asset, another portion recorded as an expense).

- Amortization of the licensed data asset attributable to AI software for which technological feasibility has not been established is recorded as R&D expense. [730-10-25-2(c), 985-20-25-1]

- Amortization of the licensed data asset attributable to a software product for which technological feasibility has been established is generally capitalizable as a production cost of the AI software, even though an allocation of such amortization may be considered an indirect, rather than direct, production cost. Indirect production costs (e.g. allocations of hardware, third-party infrastructure or developer facility costs) are capitalizable under Subtopic 985-20. [985-20-25-1, 25-3, 25-5]

Product enhancements vs maintenance

Once external-use software has been developed and released, an entity may undertake efforts to enhance the software and will almost always undertake efforts to maintain it. 'Product enhancements' extend the life or significantly improve an existing software product's marketability. Other updates (i.e. that do not meet the definition of a product enhancement) and activities to correct errors or keep the software updated with current information are considered maintenance. [985-20 Glossary]

Distinguishing product enhancements from maintenance frequently requires judgment. In the context of AI software, this includes assessing whether additional training or retraining of the software creates new functionality (i.e. new abilities) or solely maintains the software's relevancy (e.g. merely permits the software to continue to perform its existing tasks with current data). Involvement of the software development team will often be necessary to make this assessment; involving them early in the process will help to ensure appropriate conclusions are reached.

Data costs incurred for product enhancements

Under Subtopic 985-20, the accounting for product enhancement development is the same as the accounting for new software product development. Therefore, the same considerations as outlined above apply to costs of licensed data to be used in developing software product enhancements. [985-20-55-18]

Data costs incurred for maintenance

Software maintenance costs, including data costs incurred to maintain AI software, are expensed as incurred. [985-20-25-6]

📋 Internal-use software

Accounting for internal-use software development costs

When to begin and cease capitalization

Before the adoption of Accounting Standards Update (ASU) 2025-06, Targeted Improvements to the Accounting for Internal-Use Software, development costs incurred before all of the following occur are expensed as incurred: [350-40-25-12]

- the 'preliminary project stage' is complete;
- entity management authorizes (implicitly or explicitly) and commits requisite funding to the software project; and
- it is probable that the project will be completed and the software used to perform its intended function.

Under ASU 2025-06, the first criterion no longer exists and additional guidance is provided around the third. While the point in time that capitalization commences may change for some software development projects after adopting ASU 2025-06, the discussion herein about the cost accounting that applies before

and after that point are not affected by the amendments. See chapter 3A in KPMG Handbook, Software and website costs, for additional guidance on applying Subtopic 350-40 post-ASU 2025-06.

Eligible development costs incurred after the above criteria are met until the software is substantially complete (i.e. after all substantial testing is completed) are capitalized. Sections 3.2.30 (pre-ASU 2025-06) and 3A.2.30 (post-ASU 2025-06) in KPMG Handbook, Software and website costs, address what development costs are eligible for capitalization. [350-40-25-12, 25-14]

Income statement classification of such costs is not specified in US GAAP and therefore depends on the facts and circumstances. Many entities producing internal-use software for revenue-generating purposes (e.g. sale to customers via SaaS arrangements) classify such expenses as R&D, while development costs of true internal-use software (e.g. for use in the entity's selling or administrative activities) are generally classified based on the intended functional use of the software once developed.[1]

Direct vs indirect costs

Direct costs of eligible internal-use software development and cloud computing arrangement implementation activities are capitalized, while indirect costs are not. Examples of indirect costs that are generally not eligible for capitalization include allocations of (not exhaustive): [350-40-30-3, 30-5]

- lease cost for the space used by the software developers;
- depreciation of computer servers not dedicated to the software development; and
- infrastructure-as-a-service (IaaS) fees related to compute capacity not dedicated to the software development (see Questions 3.2.55 and 3A.2.210 in KPMG Handbook, Software and website costs).

Data-related infrastructure costs would generally be indirect costs unless they relate to infrastructure obtained specifically for and dedicated to data for a specific internal-use software project; for example, servers or compute capacity obtained specifically and exclusively to house data that has no alternative use to the entity other than training a specific AI internal-use software application (see Questions 3.2.57 and 3A.2.230 in KPMG Handbook, Software and website costs).

Data with other uses

Licensed data may have one or more uses to the entity other than in training a specific AI internal-use software application.[2] For example, the entity may be developing, or have a clear intention to develop, multiple applications in which it will use the data, possibly in different stages of development and with differing likelihoods of successful deployment, or the entity may have existing, already developed AI software for which the data is necessary to its maintenance.

If licensed data has other uses to the entity, it is capitalized and subsequently accounted for in accordance with Subtopic 350-30 (see *Accounting for data costs: the basics*). No portion of the licensed data asset amortization is capitalized as part of an AI internal-use software asset under Subtopic 350-40 because any such allocation would be an *indirect* cost, ineligible for capitalization.

Data with no other uses

The accounting for costs of licensed data with no use to the entity other than developing or maintaining specific internal-use AI software depends on where the software resides in the development lifecycle when those costs are incurred.

- Costs incurred before the Subtopic 350-40 capitalization criteria are met are expensed as incurred. [350-40-25-1, 25-12]

- Costs of licensed data to be used in software development (e.g. to train the AI model's core functionalities) incurred *after* the Subtopic 350-40 capitalization criteria are met, but before the

software is 'substantially complete and ready for its intended use', are capitalized as part of the AI internal-use software asset. [350-40-25-2, 25-12, 25-14]

- Costs incurred after the AI software is 'substantially complete and ready for its intended use' are expensed as incurred (unless they are capitalizable upgrade or enhancement costs – see below). [350-40-25-6]

Data licensing arrangements can span many years, with new/changed data made available throughout the data license period; therefore, data costs may be incurred under the arrangement during different phases of the software's development lifecycle. This could lead to costs incurred under a single data licensing arrangement being accounted for differently (i.e. some costs expensed as incurred, and others capitalized into the AI internal-use software asset).

Upgrades and enhancements

Upgrades and enhancements are defined in Subtopic 350-40 as changes to existing internal-use software that result in additional software functionality (i.e. the ability to perform additional tasks). Modifications that merely extend the useful or economic life of internal-use software, or increase how efficiently the software operates, but do not add to the tasks the software is able to perform are not upgrades or enhancements. [350-40-25-7]

Sections 3.2.50 and 3A.2.60 KPMG Handbook, Software and website costs, further discuss identifying upgrades and enhancements.

It will often require judgment to determine whether training or retraining AI software creates new functionality or merely maintains existing functionality (e.g. keeps the software current but does not add to the tasks the software can perform). Costs of licensed data used to upgrade or enhance internal-use AI software follow the same capitalization requirements as costs of licensed data used in new software development. Costs of licensed data used solely to maintain internal-use software are expensed as incurred.

ⓘ KPMG resources

For further guidance on the accounting for software development costs, R&D, software and SaaS revenue recognition and financial statement presentation, we recommend the following KPMG resources as a starting point.

- Handbook: Software and website costs
- Handbook: Research and development
- Handbook: Revenue for software and SaaS
- Handbook: Financial statement presentation

Contact us

Scott Muir
Partner
smuir@kpmg.com

P.K. Barot
Partner
pinakbarot@kpmg.com

---

[1] ASC paragraph 730-10-15-4(c) states that costs of developing software for use in selling or administrative activities are not R&D costs.

[2] 'Other uses' in this context means the data can be used for more than one purpose by the entity; it is not equivalent to 'alternative future use' as that term is used in Subtopic 730-10. This is because Subtopic 350-40 does not define internal-use software development as an R&D activity.

Learn about us: in | **kpmg.com**