SMPTE ENGINEERING REPORT

# Artificial Intelligence and Media

SMPTE ER 1011:2025

The home of media professionals, technologists, and engineers.

Copyright © 2025 by SMPTE® - All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, with the express written permission of the publisher.

# SMPTE ENGINEERING REPORT

## Artificial Intelligence and Media



## Table of Contents



Copyright © 2023, 2025 by THE SOCIETY OF MOTION PICTURE AND TELEVISION ENGINEERS
50 Main St., Suite 1000, White Plains, NY 10606, (914) 761-1100
Approved October 13, 2025

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

*All product names, brands, and trademarks are the property of their respective owners. Use of these does not imply endorsement.*

© SMPTE 2023, 2025 – All Rights Reserved 

# 1 Introduction

## 1.1 Executive Summary

This report is intended to provide background to media professionals on artificial intelligence (AI) and machine learning (ML). The report surveys how AI/ML are being used for media production, distribution, and consumption. It explores ethical implications around modern AI systems and provides background on current standards activities and possible opportunities for future standards. It also considers the need for datasets to help facilitate research and development of future media-related applications.

## 1.2 Revision Notes

This report is a revision of ER 1010:2023, and includes the following changes:

- Adds information on open source AI
- Adds information on multimodal AI
- Updates the models mentioned with the current state of the art
- Adds more information about agentic AI
- Adds information about protocols such as Model Context Protocol and agent-to-agent protocols that are used in building AI systems
- Adds information about AI model risks and risk management
- Adds a section on security considerations for AI systems
- Updates standards landscape information

## 1.3 Scope

This report includes high-level technical background on AI/ML algorithms and systems. It discusses how AI/ML technologies are being used to generate media and automate routine or mundane tasks in the production of media. It discusses security and ethical implications of AI/ML systems. The report surveys standards relevant to the use of AI/ML in media and discusses opportunities for future standards work. It highlights the importance of data and how collaboration could help in the creation of new datasets.

# 2 Overview of Machine Learning and Open Source

## 2.1 Machine Learning

Initial attempts at defining and building artificial intelligence, from the 1950s to the 1970s, focused for the most part on the formation and manipulation of knowledge. This was—and still is—an optimal place to start. After all, acquiring, growing and integrating knowledge across dimensions of the real world is the main attribute of intelligence.

But these initial efforts sidestepped the biggest challenge in knowledge formation: abstracting low-level data into symbolic representation (objects, situations, etc.) with the goal of contextually integrating these representations into knowledge that can dynamically evolve based on context. Because they lacked big data and powerful computation, AI experts skipped to the latter part of the knowledge pipeline, creating applications that would reason through logic on pre-packaged symbols and rules built through traditional if/then statements. This was the golden age of “expert systems.”


© Copyright SMPTE 2023, 2025 – All Rights Reserved

But as AI was going through its first “winter” in the 1970s and 1980s, computer scientists, empowered by a rise in availability of data and large computational power, started pointing out the limitation in this traditional approach, asking a very good question:

> “Where does the knowledge come from?”[1].

Many complex tasks cannot be fully modeled by traditional programming (if/then) statements (e.g., driving a car, recognizing objects in a photo, or predicting stock prices). Therefore, expert systems (see Figure 1) offer no practicable way to build applications to conduct such tasks.

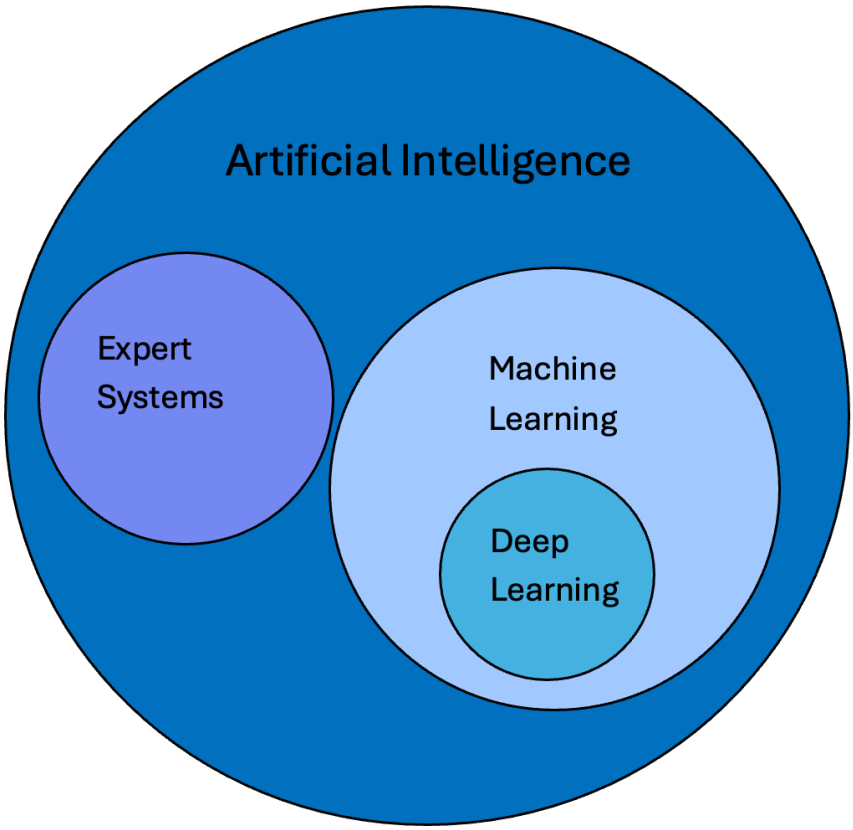


**Figure 1 — Venn diagram of expert systems and machine learning[2]**

This simple question propelled some computer scientists to almost entirely throw out the hypotheses and methods developed since 1956 and return to some of the initial AI architectures (neural networks) to start a major revolution in machine learning.

[1] Miroslav Kubat, in “An Introduction to Machine Learning”, Springer, 2017, p. ix.

[2] Note that artificial intelligence includes other domains beyond expert systems and machine learning (e.g., natural language processing)

© SMPTE 2023, 2025 – All Rights Reserved

Very simply put, "*machine learning is about extracting knowledge from data*", without having to explicitly feed a machine any information, or procedure, about this data[3].

Machine learning is also the domain of AI that has seen the most acceleration in the past ten years. Rooted in open source software, the advent of deep neural network architectures, together with the increasing availability of large (and curated) datasets and the "supercomputing for $10/hour" revolution, have all converged to produce dramatic innovation.

Today, machine learning touches almost every single aspect of our lives, from product and content recommendations to auto-tagging of pictures, digital assistants, online search, automatic voice recognition, voice synthesis, and more recently, content generation via Generative AI.

AI models generally fall into two broad categories:

- **Discriminative:** The models differentiate between data points. Those form the vast majority of supervised learning algorithms (such as classification).

- **Generative:** The models learn the underlying patterns in the training data to generate new data similar to the training data. Because they can create models from raw data, generative models are very popular in current machine learning.

There are roughly four categories of machine learning methods:

- **Supervised learning:** Algorithms that create representations and patterns from labeled data. In supervised learning, we know both the input and the output/outcome. The models are trained on each input-output pair, and with enough training examples to ensure effectiveness and accuracy.

- **Unsupervised learning:** Algorithms that recognize patterns in unlabeled data. Clustering and dimensionality reduction are both unsupervised ML methods.

- **Self-supervised learning (SSL):** Methods for processing unlabeled data to obtain representations that can be used in downstream learning tasks. The most salient feature of self-supervised learning (SSL) methods is that they do not require human annotated labels. SSL algorithms use labels automatically generated from data. For instance, a masked word in a sentence is a label and the model must predict it. To predict the masked words, SSL builds a large language model (LLM) that can then be used to perform more complex tasks than missing word prediction.

- **Reinforcement learning (RL):** Algorithms that classify input and output data according to a "reward function" acting as feedback to the agent. In reinforcement learning, artificial agents "learn" autonomously by computing all possible models between input and output data and "pruning" models based on the reward function. In his excellent book "How Smart Machines Think," Google engineer Sean Gerrish refers to this as "teaching computers by giving them treats"[4].

[3] Andreas Muller, Sarah Guido, *Introduction to Machine Learning with Python*, O'Reilly Books, 2016, p.1.

[4] Sean Gerrish, *How Smart Machines Think*, MIT Press, 2018, p.89.

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

## 2.2 Open Source Artificial Intelligence: An Expansive View

This section explores the concept of "open source" as it applies to AI, and that its meaning has expanded far beyond the traditional definition of publicly available, free-to-use computer software code. In the context of AI, open source principles now encompass data, models, and a broader range of societal and ethical considerations.

### 2.2.1 What is Open Source Software?

Most engineers are familiar with open source software (OSS), i.e., publicly available source code that is free to use, modify, and share under various licenses. The Open Source Initiative (OSI), a non-profit corporation formed to educate about, advocate for, and to build bridges among different constituencies in support of open source software, defines it more completely through its ten Open Source Definition (OSD) criteria, available at https://opensource.org/osd.

The modern OSS movement is widely credited to computer programmer and activist Richard Stallman, who launched the GNU Project in 1983[5]. Well-known examples of open source software projects include the Linux operating system and the Firefox web browser. In the media and entertainment industry, a prominent example is the OpenEXR file format and its associated code libraries, a project hosted by the Academy Software Foundation (ASWF). The ASWF is an industry-supported nonprofit organization that provides a neutral forum and infrastructure to increase the quality and quantity of contributions to open source software in content creation[6]. As of this writing, the ASWF hosts 19 open source projects and several working groups.

Licenses to use, modify, and redistribute open source software are necessary components that define the legal framework and collaboration environment. License terms run the gamut from highly permissive, i.e., placing few restrictions beyond attribution and inclusion of the original copyright notice and license, to highly restrictive, i.e., requiring derivative works and associated code to be licensed as well. Well-known examples of permissive licenses are MIT and Apache, and a more restrictive, or “copyleft” license, is GNU General Public License (GPL).[7]

The benefits of OSS are widely recognized and include transparency, vendor independence, customization, and the collective efforts of a large community. While the code is free, adopters often incur development costs for customization and integration. Potential downsides include insufficient documentation, security risks, project "forking" (splitting the code into multiple variants), and restrictive licensing. However, the widespread adoption of OSS across industries demonstrates that the benefits generally outweigh the downsides.

[5] https://www.gnu.org/gnu/thegnuproject.en.html

[6] https://www.aswf.io/

[7] A non-exhaustive list of open source licenses may be found here: https://opensource.org/licenses

© SMPTE 2023, 2025 – All Rights Reserved

### 2.2.2 Open Source Software in Machine Learning

While the open source movement gained traction with the GNU Project, the practice of sharing software in research dates back to the 1950s, a hallmark of academic and scientific collaboration[8]. Building on early AI development, including work at Stanford University's AI Lab, Mark Kantrowitz established the CMU Artificial Intelligence Repository at Carnegie Mellon University in 1993. This repository was created to "collect free software and materials of general interest to AI researchers" and has since become a vital archive of research papers, code, and other artifacts[9].

Today, the landscape of open source machine learning tools is vast. Well-known examples include Google's TensorFlow[10], the Linux Foundation's PyTorch[11] (originally developed by Meta), and a wide variety of tools and resources from Hugging Face[12] and NVIDIA[13]. The Journal of Machine Learning also maintains a list of dozens of open source projects for research and application needs[14].

### 2.2.3 From Open Source Software to Open Source AI

The expansion of "openness" in AI systems has led to a more comprehensive definition. In 2024, the Open Source Initiative published its "Open Source AI Definition" as version 1.0, acknowledging the evolving nature of AI. This definition extends beyond source code to include other essential components:

> An open source AI is an AI system made available under terms and in a way that grants the freedoms to:
>
> - Use the system for any purpose and without having to ask for permission.
> - Study how the system works and inspect its components.
> - Modify the system for any purpose, including to change its output.
> - Share the system for others to use with or without modifications, for any purpose.
>
> These freedoms apply both to a fully functional system and to discrete elements of a system. A precondition to exercising these freedoms is to have access to the preferred form to make modifications to the system[15].

According to this definition, a truly open source AI system must share more than just the software source code. It also requires access to the legally shareable training data and code, the resultant model "weights" (the numerical parameters learned during training), and sufficient documentation for developers to understand and build upon the system.

---

[8] https://thelinuxcode.com/a-brief-history-of-open-source/

[9] https://www.cs.cmu.edu/Groups/AI/0.html

[10] https://www.tensorflow.org/

[11] https://pytorch.org/

[12] https://huggingface.co/

[13] https://developer.nvidia.com/open-source?sortBy=open_source_projects%2Fsort%2Ftitle%3Aasc

[14] https://www.jmlr.org/mloss/

[15] https://opensource.org/ai/open-source-ai-definition


© Copyright SMPTE 2023, 2025 – All Rights Reserved

This definition clarifies the distinction from "open-weight" AI systems (described in Section 8.2.2). While open-weight models make the trained model weights publicly available, they do not share the underlying training data or the source code used to produce the weights. Therefore, they do not meet the OSI's 1.0 definition of open source AI. The OSI provides a non-exhaustive list of AI systems that comply with their definition[16].

Licensing of AI systems and components extends beyond the source code license regimes described above. The concept of Responsible Artificial Intelligence Licenses (RAIL) emerged from the responsible and ethical AI community with the intention of providing a framework for restrictions on how open source AI code, data, and applications (and their derivatives) might be used. The Responsible AI Initiative hosts a website[17] with extensive information on this topic.

For developers seeking to evaluate the "openness" of a model, the Model Openness Framework (MOF) is a useful resource. Developed by the LFAI+Data Foundation and the Generative AI Commons (both projects of the Linux Foundation), the MOF provides a set of guidelines and a tool that generates a numerical score indicating the degree to which a model aligns with its principles[18].

### 2.2.4 Beyond the Code: Public Policy and Human Impact

The word "software" was deliberately omitted from this section's title to reflect the expansive nature of "openness" in the context of artificial intelligence. According to a 2025 Black Duck report on OSS security, over 97% of commercial software projects scanned contained open source components, underscoring its ubiquity[19].

Now that open source principles extend to training data and model weights, engineers building AI systems must navigate new complexities, including data licensing and governance. The extraordinary range of AI applications and their human impact raise concerns that go far beyond traditional code-sharing and licensing. Terms such as public policy, sovereignty, ethics, regulation, and “open-washing”[20]—not historically associated with the OSS movement—are now central to the discussion.

While proprietary AI models are not inherently bad, statutory requirements like the European Artificial Intelligence Act and other regulatory actions discussed later in this report require engineers to consider a broader range of issues when adopting or contributing to open source AI projects. Organizations such as the OSI, the Linux Foundation (LF), and the ASWF are excellent resources for engineers to become informed about these issues and to be supported in their open source AI journey.

---

[16] https://opensource.org/ai

[17] https://www.licenses.ai

[18] https://isitopen.ai/

[19] https://www.blackduck.com/resources/analyst-reports/open-source-security-risk-analysis.html

[20] ”Open-washing” is the practice of presenting a technology as open when it is, in fact, not.

© SMPTE 2023, 2025 – All Rights Reserved

# 3 Deep Learning

In the strict sense, deep learning is not related to the training process but to its fundamental architecture. The deep learning models can be trained using supervised, self-supervised, or unsupervised algorithms. Deep learning models use multiple layers to progressively extract higher-level features from the raw input and do not need features generated by external processing. Some aspects of deep learning models are bio-inspired. Indeed, artificial neural networks (ANNs) are AI models inspired by biological neural networks. At the core of these systems are interconnected units or nodes, called artificial neurons, each of which produces a real-valued output. Stacking layers of these artificial neurons, ANNs can perform complex tasks on unstructured signals like audio or video as humans do.

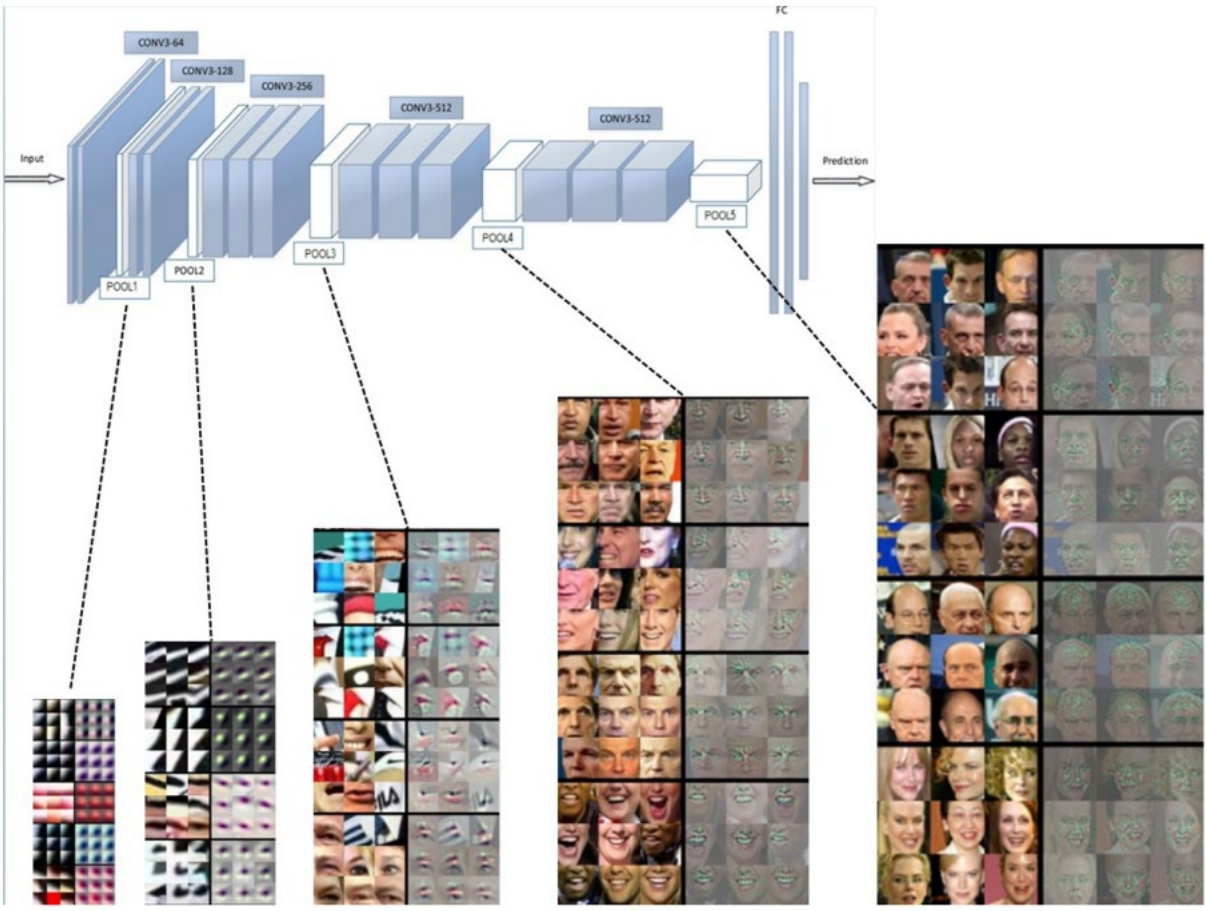


**Figure 2 — Layers and generated features of a deep neural network for facial recognition, source[21]. At the left of the layers, are the stimuli (a fraction of the inputs), and at the right, the outputs to these stimuli.**

[21] Masi, Iacopo, Yue Wu, Tal Hassner, and Prem Natarajan. "Deep face recognition: A survey." In *2018 31st SIBGRAPI conference on graphics, patterns and images (SIBGRAPI)*, pp. 471-478. IEEE, 2018.


© Copyright SMPTE 2023, 2025 – All Rights Reserved

Deep learning models are based on ANNs, but with several 'hidden' layers between the input and output layers. These intermediate layers transform the input data into a form that the output layer can use for prediction. Hidden layers in deep learning refer to the extended layers through which data is transformed, enabling the model to learn and extract complex patterns via a hierarchy of progressively complex features. More layers are used to build a representation of the input signal that will be well suited to match the input signal and the labels. Figure 2 illustrates the functionalities of the layers for facial recognition systems. In this case, it is interesting to notice that the first layers are very similar to filters created by humans for image processing, 30 years ago with a mathematical approach. The higher layers learn more complex features that are humanly understandable.

The impact of deep learning is significant and widespread, pushing boundaries in a variety of fields. It has contributed to breakthroughs in areas such as computer vision, natural language processing, speech and audio recognition, and even bioinformatics. In the field of computer vision, convolutional neural networks, a specific class of deep learning models, have succeeded in recognizing and identifying faces, everyday objects, road signs and much more.

# 4 Supervised Learning

## 4.1 Overview

Supervised learning is a robust and prevalent type of machine learning, in which algorithms predict a pre-labeled outcome from a given input.

For example, if we were to build an application that can distinguish dogs from cats in pictures, then we might feed a supervised learning algorithm hundreds of thousands (often millions) of pictures of dogs and cats labeled “dog” and “cat.”

However, to be powerful enough to recognize many different types of dogs and cats, the learning algorithm must be fed training data with a wide range of diverse samples. This highlights the main constraint of supervised learning: it is entirely reliant on the size, diversity and structure of the training data.

There are two major types of supervised learning methods: classification and regression.

## 4.2 Classification

As its name implies, classification seeks to predict the class of the input data among a set of labels that are explicit in the data. For example, “spam” vs “not-spam” are two classes; dogs, cats, elephants and penguins are four classes; etc. Classification can be binary (between a set of two classes) or multiclass (among more than two classes).

## 4.3 Regression

Regression seeks to predict not a class but a continuous numerical value, so it is restricted to numerical data. It can predict the statistical relationship between a set of numerical variables. One of the most commonly featured examples of regression problems is predicting home values based on various numerical features of the home.

© SMPTE 2023, 2025 – All Rights Reserved

# 5 Unsupervised Learning

## 5.1 Overview

Unsupervised learning is an important part of a more exploratory approach in data science, and can help with large and highly dimensional datasets, but is hard to evaluate because there is no single correct answer (“ground truth”); however, it provides an effective set of tools to surface the latent variables in a highly dimensional dataset.

Unsupervised learning falls mainly into two categories and related use cases: dimensionality reduction and clustering.

## 5.2 Dimensionality Reduction

As its name implies, dimensionality reduction is a set of techniques that learn fundamental patterns in data to compress it into a simpler, more compact representation maximizing the information density.

## 5.3 Clustering

Clustering groups together data points that are similar to one another. A clustering algorithm detects similarity and classifies the input data into a number of distinct clusters. Algorithms may vary in how they classify the input data, which criteria they optimize for, and which assumptions are made in the process.

# 6 Self-supervised Learning

Self-supervised learning has an advantage over supervised learning in that it does not require labelled data, but is different from unsupervised learning, because labels (the data to predict) are extracted from training data itself. For example, in text, the data driving the training can be hidden words to be guessed, or in images, it can be masked pixels to be filled.

Machine learning models can be either discriminative or generative, as explained in the introduction, and the recent generative models that have been a breakthrough in the field of AI are trained using a self-supervised approach. As there is no need to annotate the data for training, self-supervised learning allows generative models to exploit a diverse and massive amount of data. This explosion of the size of the training set suggests that some models might be being trained using unchecked data that may be biased or of poor quality.

The basic idea behind a generative model is to take a large number of examples from the training set and to learn the probability distribution that generates those training examples. As an example, the predominant LLMs use autoregressive models, a type of probabilistic model that predicts the probability of a word (or a token) in a sequence based on the preceding ones.

The dream of AI researchers is to be able to generate knowledge about the world as a whole. And as humans know, the world is messy. So generative machine learning algorithms need to learn from data that is sometimes large, but most often sparse, unstructured, and biased. Section 8 on generative AI explains the principle and application of some generative self-supervised models.


© Copyright SMPTE 2023, 2025 – All Rights Reserved

# 7 Reinforcement Learning

In their monumental “Reinforcement Learning: An Introduction,” Richard Sutton and Andrew Barto define Reinforcement Learning as follows:

> “Reinforcement Learning is learning what to do—how to map situations to actions—so as to maximize a numerical reward signal. The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them.”[22]


**Figure 3 — Principle of reinforcement learning**

In Reinforcement Learning (see Figure 3) the environment first presents an initial state to the agent, which causes it to generate an action. This action results in a reward for the agent, which, when processed by the reward function, leads to the creation of another action aimed at maximizing the reward. After many rounds of this cycle, which is essentially a process of trial and error, the agent learns to identify the optimal sequence of actions, or the policy, that leads to the maximization of rewards in the environment.

In summary, Reinforcement Learning maps an agent's state to the best (probabilistically estimated) action using the reward function, which evaluates the reinforcement signal of each action so that the next action further maximizes the reward and gets closer to the agent's goal in the environment. It is believed that with enough training through RL, an agent can surpass even human intelligence in certain domains, including quite complex ones. Video games are a major area where computational agents are already approaching, and sometimes surpassing, human-level intelligence.

[22] Richard Sutton, Andrew Barto, *Reinforcement Learning: An Introduction*, 2nd Edition, Cambridge, MIT Press, 2018, p.1.

© SMPTE 2023, 2025 – All Rights Reserved

# 8 Generative AI

## 8.1 Overview

Generative AI is a type of machine learning in which a model is trained to generate new data. The goal is to create a representation of the data that will serve as an engine to generate other data. For example, if the data is text, a language model will be generated. Unlike discriminative models that predict labels based on input features, generative models use those features to understand and replicate the data distribution. Generative models have myriad applications, including creating realistic images, composing music, and writing text. Notable examples of generative models include Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), diffusion models such as OpenAI DALL-E 2[23] or Midjourney[24] for image generation, and transformer-based models like chat-GPT[25] (Generative Pretrained Transformer) for text generation.

Multimodal generative AI models extend these capabilities by processing and integrating information from multiple data types, such as text, images, audio and video. These models can perform cross-modal tasks, such as generating images from text descriptions or creating captions for visual content, enabling more comprehensive and nuanced AI applications.

One of the most interesting aspects of generative AI is its potential to produce new and creative solutions to problems that are difficult or impossible to solve using traditional rule-based programming methods or other machine learning approaches. In the media domain, Large Generative AI Models (LGAIM) are used to create images, video clips, texts, music and creative works. Table 1 provides an overview of the popular LGAIM models.

**Table 1 — Examples and capabilities of selected Large Generative AI Models**

| Data generated | Training | Example | Capabilities |
|---|---|---|---|
| Text | Words or word tokens | GPT-4, Claude 3.5, Gemini 1.5, Llama3.1, FALCON, Mistral large 2 | NLP, Natural Language Generation, Translation |
| Images | Images with text caption | DALL-E 3, Midjourney, Stable Diffusion, Adobe Firefly, ImageFX | Images |
| Music | Audio waveforms of recorded music with text annotations | MusicLM, MusicGen, Stable Audio, AIVA | Music from text |
| Video | Annotated video | Synthesia, Pika Lab, Colossyan, Fliki, Stable Diffusion Video | Video clips |

[23] *DALL·E 2 is an AI system that can create realistic images and art from a description in natural language*

[24] *Midjourney generates images from natural language descriptions; the tool is designed by an independent research lab Midjourney, Inc.*

[25] *chat-GPT* ChatGPT is a large language model-based chatbot developed by OpenAI that allows users to refine and direct a conversation and perform natural language processing tasks such as summarizing or tagging.

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

| Data generated | Training | Example | Capabilities |
|---|---|---|---|
| Text | Text and image | CLIP, LLaVA, Vid2Seq | Image and video annotation, Visual understanding via text |
| 3D models | Text and 3D outputs | Point-E, FlexiCubes | 3D meshes, point clouds |
| Text | Text, images, videos, audio | GPT-4o, Flamingo, Kosmos-2 | Automated image and video description<br>Multimodal content analysis |

## 8.2 Large Language Models

### 8.2.1 Overview

Large Language Models (LLMs) are the foundation of generative AI systems capable of producing coherent, human-like text. Trained using a self-supervised objective to predict the next token, typically sub-word units, LLMs ingest hundreds of billions of tokens and learn to model long-range dependencies using transformer self-attention. This architecture enables them to generate fluent, context-aware text across languages and domains.

### 8.2.2 Openness and Access Models

LLMs differ in their openness. ChatGPT, for example, is accessible only via an API or user interface and remains fully proprietary. In contrast, Meta's LLaMA 3 is released as an open-weight[26] model: its pre-trained weights are available for commercial use by organizations with fewer than 700 million monthly active users. However, it lacks training code and datasets and thus is not fully open source.

Some models meet stricter open source criteria by releasing not only the model weights but also the training code, datasets (or their detailed descriptions), and alignment tools under permissive licenses. Alignment tools refer to the components used to fine-tune the model's behavior according to human preferences or safety objectives.

As an example, Dolly 2.0 (Databricks, 2023) is a 12B-parameter model released under Apache 2.0 that includes not only the model weights but also the complete fine-tuning pipeline and a 15k-example instruction dataset, enabling full transparency and reproducibility.

### 8.2.3 Scaling Laws and Computational Efficiency

Performance improvements in LLMs follow well-established scaling laws, which show that increasing model size, data, and compute leads to predictable gains. These laws were first described by Kaplan[27] and later refined by Hoffmann[28] in the Chinchilla study. The latter demonstrated that Chinchilla, a 70B model trained on 1.4 trillion tokens, outperformed a 175B model trained on less data, highlighting the value of compute/data balance.

---

[26] "Open-weight" refers to models that provide pre-trained weights for inference or fine-tuning but withhold the full training code and datasets. These models may also be subject to license-based usage restrictions.

[27] Kaplan et al., "Scaling Laws for Neural Language Models," 2020. https://arxiv.org/abs/2001.08361

[28] Hoffmann et al., "Training Compute-Optimal Large Language Models," 2022. https://arxiv.org/abs/2203.15556

© SMPTE 2023, 2025 – All Rights Reserved

Training frontier LLMs is highly resource-intensive, often requiring thousands of GPU-days and datasets exceeding one trillion tokens. To improve scalability and reduce the cost of serving these models in production, developers frequently publish size variants (e.g., 7B, 13B, 70B) that trade peak accuracy for lower inference latency and compute requirements. Further gains come from Mixture-of-Experts (MoE) architectures, which reduce computational load by activating only a subset of model parameters per input. For instance, Mixtral 8×7B (Mistral AI, 2023) routes each token through 2 of 8 expert networks, while DeepSeek-V2 (DeepSeek AI, 2024) uses 2 of 64, dramatically reducing floating-point operations while maintaining high performance. This modular design not only improves efficiency but also lays the groundwork for agent-like architectures, where specialized components collaborate across tasks in adaptive, intelligent workflows.

### 8.2.4 Reasoning and Learning Strategies

LLMs are not only fluent, but they increasingly demonstrate emergent reasoning abilities, including logical deduction, and instruction following. These behaviors are supported by two key mechanisms:

- Inference-time prompting, such as Chain-of-Thought (CoT), which guides the model to produce intermediate reasoning steps, improving transparency and accuracy.
- Training-time supervision, including model scaling, instruction tuning, CoT datasets, and reinforcement learning with human feedback (RLHF), which helps models internalize multi-step strategies.

State-of-the-art models such as GPT-4o and Claude 3 Opus combine fluent generation with step-by-step reasoning, narrowing the gap between implicit and explicit cognition. Still, Chain-of-Thought prompting remains valuable when auditability is required.

### 8.2.5 Tool Use and Agentic Capabilities

Evolution in LLM design involves interaction with external systems through tool use and retrieval. This includes actions such as fetching real-time information, invoking APIs, or performing external computations.

Toolformer[29] enables tool use by training a model (via self-supervised data augmentation) to insert tool/API calls during text generation. It operates in a single forward pass without feedback.

In contrast, ReAct[30] models interleave reasoning and acting: the model generates a thought, executes an action (e.g., search), observes the result, and continues. This feedback loop supports dynamic, multi-step decision-making, making ReAct more suitable for interactive or agent-based tasks. Together, these approaches extend LLMs beyond reasoning toward real-world functionality and agentic behavior.

[29] Schick et al., “Toolformer: Language Models Can Teach Themselves to Use Tools,” 2023. https://arxiv.org/abs/2302.04761

[30] Yao et al., “ReAct: Synergizing Reasoning and Acting in Language Models,” 2022. https://arxiv.org/abs/2210.03629


© Copyright SMPTE 2023, 2025 – All Rights Reserved

### 8.2.6 Multimodal Models

Building upon these advances, multimodal models represent a new frontier in generative AI by integrating and jointly processing information from multiple data modalities, such as text, images, audio, and video. These models are designed to understand and generate content across different formats simultaneously, enabling tasks that require cross-modal reasoning. For instance, OpenAI's GPT-4o and DeepMind's Gemini integrate vision and language capabilities, allowing the system to interpret an image and answer questions about it in natural language, or generate descriptive captions. Architecturally, multimodal models often extend transformer-based LLMs with additional encoders or adapters to process non-text inputs and align latent representations across modalities. This allows them to perform complex tasks such as visual question answering and image generation from text prompts. Multimodal AI systems open new possibilities for applications in media, accessibility, and human-computer interaction, marking a shift toward more general-purpose and human-like AI agents.

At the same time, generative models focused on specific modalities, particularly vision, have also seen major advances. In computer vision, models such as GANs, VAEs, and diffusion models are widely used to generate realistic images, videos, and more. These models often serve as components in larger multimodal systems or as standalone tools for synthetic content creation.

## 8.3 Variational Auto-encoders

A Variational Auto Encoder (VAE, see Figure 4) is a type of Auto Encoder used for learning efficient representation of input data for the purpose of generating new data. It consists of two main parts: an encoder, which compresses the input into a latent-space representation, and a decoder, which reconstructs the input data from this compressed representation. The latent space contains the minimal information required for data representation, which can be used to generate new data. VAEs are trained to minimize the reconstruction error with the constraint of providing the generation capability; this is the key difference compared to classical Auto Encoders. It aims to produce outputs that closely match the original inputs from the latent space. For the generation of new outputs, a random noise that is consistent with the learned statistical properties is created in the latent space and filtered by the decoder.

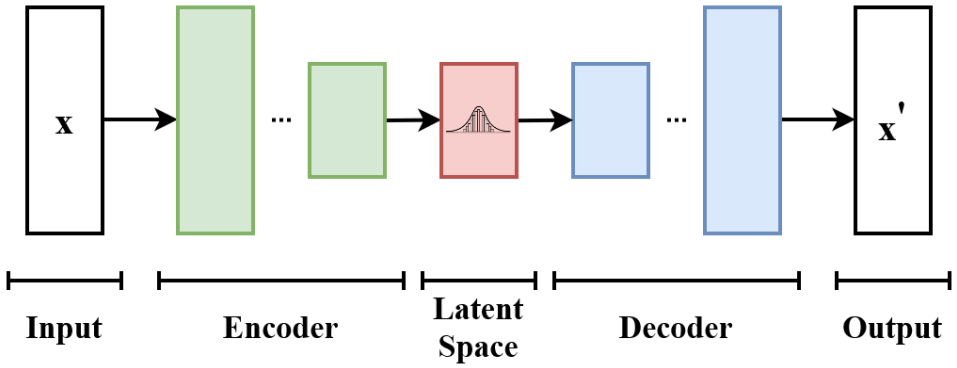


**Figure 4 — Principle of VAE[31]**

VAEs are widely used in tasks like noise reduction, feature extraction, anomaly detection, and image or sound generation.

[31] https://en.wikipedia.org/wiki/Variational_autoencoder

© SMPTE 2023, 2025 – All Rights Reserved

## 8.4 Generative Adversarial Networks

A Generative Adversarial Network (GAN) is a class of machine learning systems invented by a Google AI research team led by Ian Goodfellow[32] in 2014. A GAN consists of two neural networks, a generator, and a discriminator, which are trained together (see Figure 5). The goal of the generator is to produce data that is indistinguishable from real data, while the goal of the discriminator is to distinguish between real and fake data. Through this adversarial process, both networks improve over time, with the generator producing increasingly realistic data.

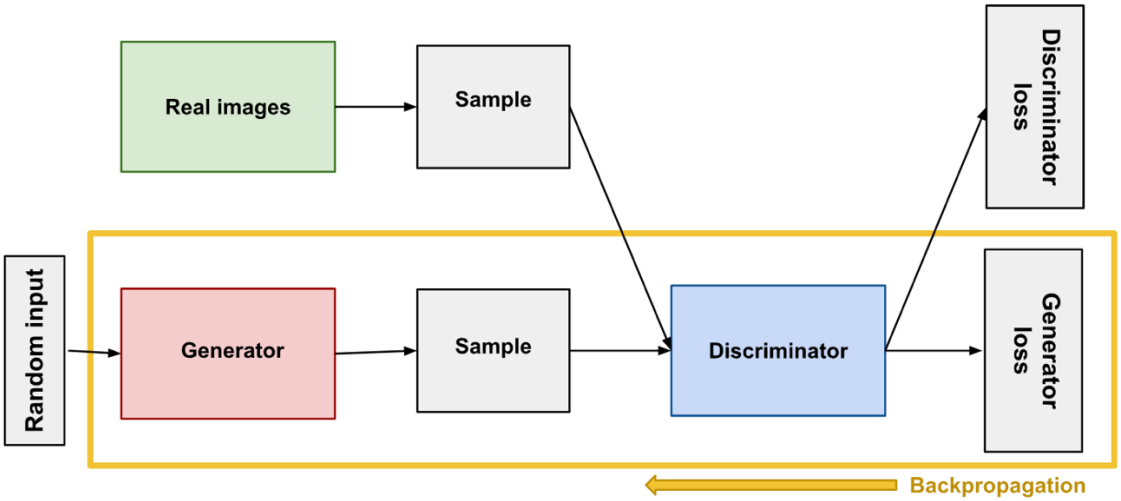


**Figure 5 — Training principle of GANs[33]**

GANs can be used for content creation:

- to generate new data for training AI models
- to generate and transform images
- for voice cloning
- to generate deepfake videos
- to generate paintings from photos
- to animate photos from videos

[32] Ian J. Goodfellow is an American computer scientist, engineer, and executive, most noted for his work on artificial neural networks and deep learning.

[33] https://developers.google.com/machine-learning/gan/gan_structure


© Copyright SMPTE 2023, 2025 – All Rights Reserved

One popular application of GANs is the generation of realistic human faces as shown in Figure 6.

**Figure 6 — Fake face generated by a GAN on https://thispersondoesnotexist.com/**

© SMPTE 2023, 2025 – All Rights Reserved

## 8.5 Diffusion Models

Diffusion models are generative models that use a two-step process to train (see Figure 7). A standard diffusion model has two major domains of processes: forward diffusion and reverse diffusion. In a forward diffusion stage, the image is corrupted by gradually introducing Gaussian noise until the image becomes complete random noise.[34] In the reverse process, a series of Markov chains are used to recover the data from the Gaussian noise by gradually removing the predicted noise at each time step inference.[35]

These models are highly computationally demanding, and training requires a very large memory, which makes it impossible for most practitioners to even attempt the method. Diffusion Models have recently shown a remarkable performance in Image Generation tasks and have superseded the performance of GANs and VAEs in this field. Diffusion models are considered to be foundation models, which are those that can be adapted to a wide range of downstream tasks. Foundation models are large-scale AI models that are trained in a self-supervised manner on large amounts of unlabeled data. Among the most popular examples of foundation models are diffusion models, GANs, LLMs and VAEs, which power well-known tools such as ChatGPT, DALLE-2, Segment Anything and BERT. Researchers have employed Distribution Matching Distillation (DMD) to reduce the computation required for some diffusion models.[36]

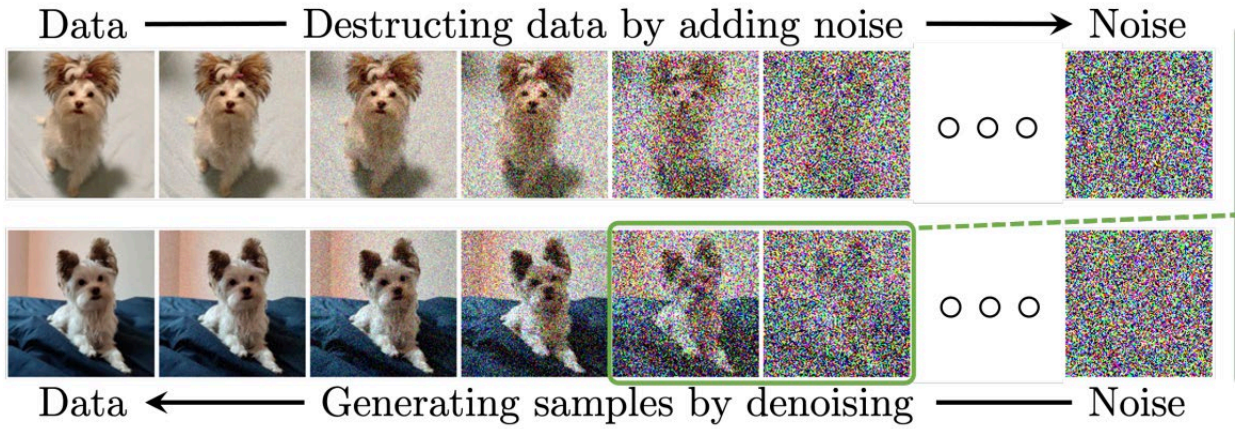


**Figure 7 — General principle of diffusion models[34]**

[34] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in neural information processing systems* 33 (2020): 6840-6851.

[35] Ian J. Goodfellow is an American computer scientist, engineer, and executive, most noted for his work on artificial neural networks and deep learning.

[36] Yin, T., et. al. Improved Distribution Matching Distillation for Fast Image Synthesis, 2024. https://arxiv.org/abs/2405.14867v2


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 8.6 LLM Benchmarking

Large Language Models (LLMs) are a class of LGAIMs that are trained to generate text. In terms of training strategy, self-supervised learning has been a key factor in the success of the LGAIMS. LLMs are trained predominantly on large amounts of raw text data using a language modelling task. In the case of an auto-regressive model, this task aims to predict the next most likely word in a sequence based on the previous words. By training on this task, the model learns to grasp the complex statistical structure of natural language, including syntax and semantics, allowing it to produce high-quality text. This is an active area of research, but off-the-shelf products that leverage LLMs are now available to end users. For example, OpenAI ChatGPT is a popular product that was built on top of GPT-3.5, an LLM built and commercialized by OpenAI. LLMs are the foundation for almost all major language technologies, but their capabilities, limitations, and risks are not well understood.

As the training of the LLM is performed on a pretext task (such as the prediction of the next words), their evaluation becomes a complex problem. The performance on this pretext task cannot necessarily predict the performance on other tasks such as content annotation, summarization, natural language generation.

In recent years, several benchmarks have been developed to assess the performance of LLMs. But there is no consensus, and the benchmarks vary in their scope and evaluation criteria, making it difficult to compare the results. Several benchmarking frameworks are available in the research area, including Pile[37], SuperGLUE[38], MMLU[39] and HELM[40].

One of the most used frameworks, Holistic Evaluation of Language Models (HELM), is a comprehensive framework for improving the transparency of language models. It involves the design of a taxonomy to describe the vast space of potential test scenarios and metrics of interest. A broad subset of cases is selected based on coverage and feasibility, and a multi-metric approach measures performance, accuracy, calibration, robustness, fairness, bias, toxicity, and efficiency.

To overcome the problem of increasing complexity and number of tasks to be evaluated, another approach to benchmarking is the minimalist approach proposed in the LMentry[41] framework. This involves assessing the basics of performance before extrapolating to more complex tasks, using simple language tasks that a primary school child might answer.

In short, the search for a comprehensive benchmark for evaluating language models in all aspects of real-life language use is still ongoing. Some opportunities for standards in this area are outlined later in this report.

---

[37] Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., Nabeshima, N. and Presser, S., 2020. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*.

[38] Wang, A., Pruksachatkun, Y., Nangia, N., Singh, A., Michael, J., Hill, F., Levy, O. and Bowman, S., 2019. Superglue: A stickier benchmark for general-purpose language understanding systems. *Advances in neural information processing systems*, *32*.

[39] Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D. and Steinhardt, J., 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

[40] Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., Zhang, Y., Narayanan, D., Wu, Y., Kumar, A. and Newman, B., 2022. Holistic Evaluation of Language Models.(2022). https://arxiv.org/abs/2211.09110.

[41] Efrat, A., Honovich, O. and Levy, O., 2022. LMentry: A language model benchmark of elementary language tasks. *arXiv preprint arXiv:2211.02069*.

© SMPTE 2023, 2025 – All Rights Reserved

# 9 MCP and A2A: Complementary Protocols for AI Interoperability

## 9.1 MCP and A2A Usage

Model context protocol (MCP) is a de facto framework for connecting AI applications to external systems. It was introduced by Anthropic in 2024 and has been adopted by other AI developers.

While MCP focuses on how intelligent agents interact with tools and data sources within their operational environment, Google's A2A (Agent-to-Agent) protocol addresses the complementary challenge of enabling independent agents, possibly from different vendors or organizations, to find each other, advertise their capabilities, exchange tasks, and collaborate securely (see Figure 8). A2A defines an Agent Card, which is a published capability description. It supports flexible communication methods, ranging from simple, web-based messaging to high-performance data exchange. A2A also enables agents to share progress updates in real time rather than waiting for complete results.

In a media workflow, this complementary architecture enables sophisticated orchestration. For example, an editing assistant agent could use MCP to query a media asset management system and access editing tools. Then, it could leverage A2A to delegate a rights-clearance task to a specialized compliance agent and coordinate publishing timelines with a scheduling agent, all without hard-coded integrations. The distinction is architectural: MCP standardizes agent-to-tool communication for capability enhancement, and A2A standardizes agent-to-agent communication for collaborative workflows.

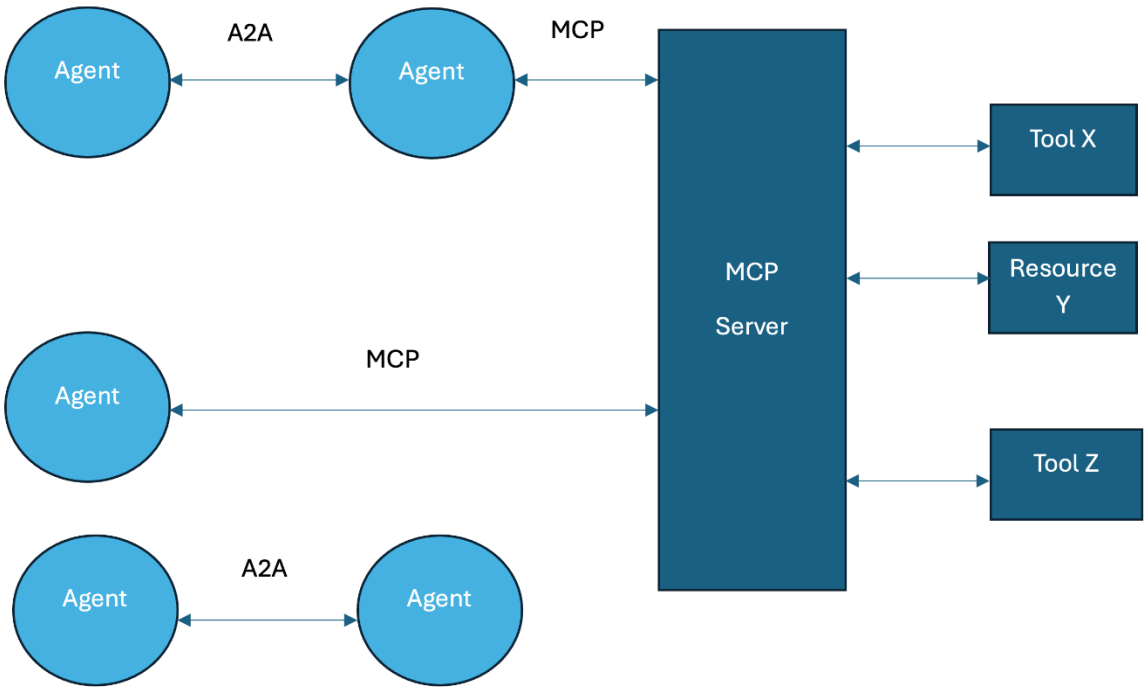


**Figure 8 — Illustration of how agentic applications use A2A and MCP**


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 9.2 Agents and Agentic Workflows

Agents are software entities with some level of agency that autonomously or semi-autonomously perform tasks on behalf of a user to pursue the user's goals. Agents make use of advanced reasoning and memory capabilities and make use of other systems such as LLMs.

Currently, the agency of agents lies somewhere between deterministic chatbots and human agency, as shown in Figure 9 from 2025 Top Strategic Technology Trends, Gartner, 2025.

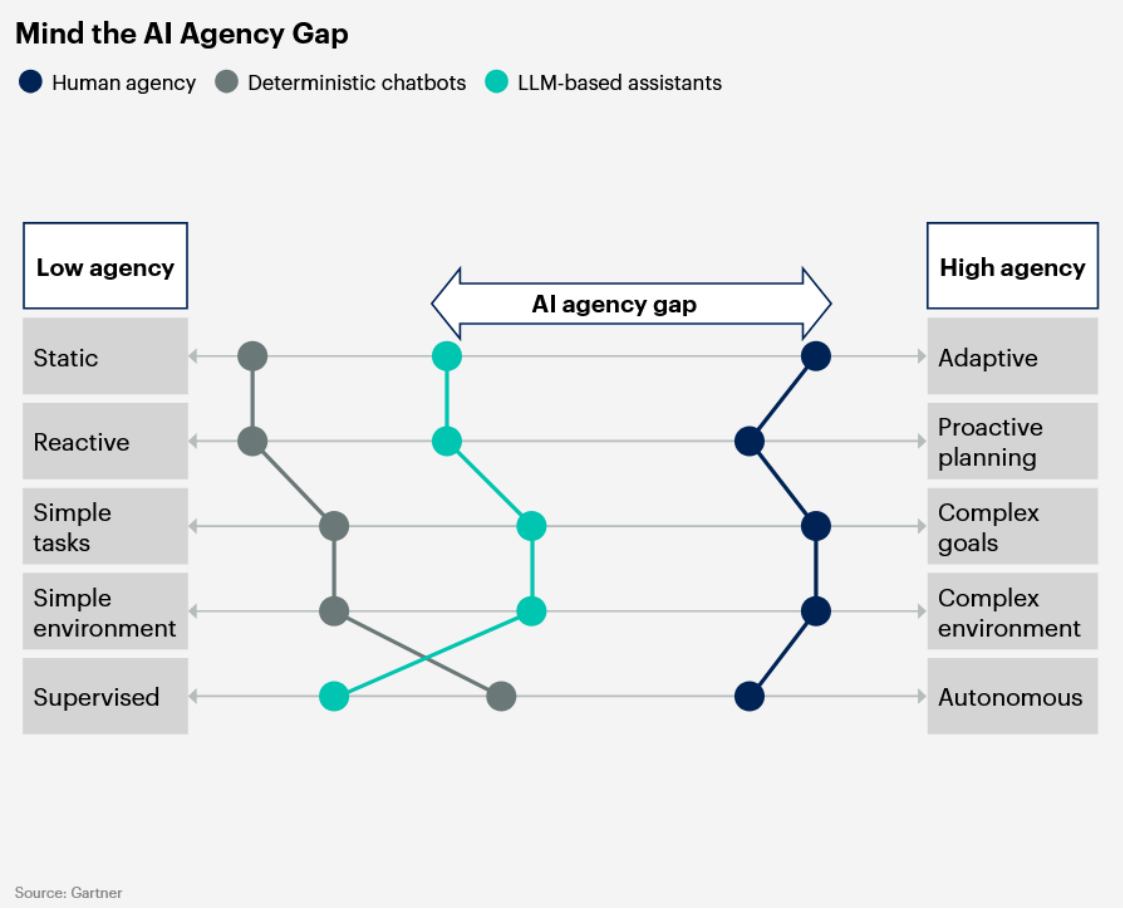


**Figure 9 — AI agency gap (from Gartner 2025)**

However, this diagram represents agency in mid-2025, and it will rapidly become out of date such is the pace of AI development. The company Protect AI presented this progression at Black Hat USA 2025.

1. LLM Application
2. Function Tool Agent
3. Tool-abstract (MCP) Agent ← We are here
4. Auto tool discovery Agent
5. Self-modifying Agents
6. General Agents

We are at step 3, so there is a long way to go in terms of functionality but not necessarily in time.

© SMPTE 2023, 2025 – All Rights Reserved

# 10 Security in AI systems

This section is a short review of the security landscape for AI systems in the arena of media creation other than the security of the execution infrastructure (e.g., cloud servers and storage) and APIs where the risks are common to non-AI systems.

The risks associated with AI systems include intellectual property (IP) protection, non-compliance with organizational requirements, and threats to the security of the workflows that incorporate AI.

AI systems can be attacked in a variety of ways, including unintentionally, and securing these systems means solving new and more complex security problems.

## 10.1 Terminology

In this section, an **AI system** is an AI model or composition of AI models or application/system/service that uses such model or models. This includes AI agents.

A **compromised** AI system is one whose behavior has been modified by a bad actor or unintentionally by a good actor such that, for example, it produces incorrect or inappropriate output, leaks the Intellectual Property (IP) in the model, causes media production workflows to malfunction, and so on.

We define **IP compromise** to mean a compromise of the IP associated with or built into the AI system. For example, the exfiltration of IP including IP in the training data and IP in the model, and the unauthorized introduction of foreign IP[42]. The exfiltration might be direct or by inference and is not confined to bulk exfiltration.

The word **Participant** is defined by the MovieLabs Ontology for Media Creation[43] to be the entities (people, organizations, or services) that are responsible for the production of a Creative Work.

In the field of AI, a **Guardrail** is a mechanism designed to ensure that AI works within a set of defined parameters including **Alignment** with organization policies and norms. Guardrails may be built into an AI system by the developers or bolted on by the users of the AI system.

## 10.2 Compliant Use and IP Protection

AI use must comply with contractual obligations (licensing agreements, labor contracts, etc.), territorial regulations, corporate policy, etc. Compliant use also means responsible and ethical use meaning, for example, AI generated imagery should not contain offensive or inappropriate elements.

The risks to IP protection include exfiltration of the IP in an AI model, the introduction of foreign IP into an AI model, and the leakage of user interactions.

Risk management in this context requires mechanisms for risk assessment that might include quantifying the trade-off between risk and capability. The trade-off is between how much autonomy an AI system has to accomplish the task, for example, in assembling news stories, and the risks associated with the output being unacceptable in some way, for example, serious inaccuracies in a news story.

---

[42] For example, the unauthorized introduction of unlicensed IP into an AI system may threaten ownership of the IP in the AI model.

[43] Ontology for Media Creation, v2.6, https://movielabs.com/production-technology/ontology-for-media-creation/


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 10.3 Attacks on AI Systems

There are many ways that an AI model can be compromised. Here, we describe two classes of attack: data poisoning and jailbreaking.

While it is more than conceivable that an AI model might poison or jailbreak itself, AI system self-harm and anti-social behavior are outside the scope of this discussion.

### 10.3.1 Data Poisoning

Data poisoning is the intentional or unintentional compromise of the training dataset used by the AI model. In this definition, we exclude data poisoning of inference-time inputs which can be classified as jailbreaking (see below) and poisoning of interference outputs which is not specific to AI systems and is an API/system security issue. We include unintentional compromise since a well-meaning use of the system can cause data poisoning by, for example, introduction of incorrect training data.

Poisoning can result in the compromise of the AI system's IP and undesirable system performance, including output that violates definitions of compliant use.

### 10.3.2 Jailbreaking

Jailbreaking is the intentional or unintentional violation or bypassing of guardrails put in place to regulate the behavior of the AI system, for example to ensure compliant use of the model. Some of these guardrails may have been put in place by the developers, and some may originate from an AI system, such as an AI agent that relies on the AI model.

Prompt injection is an attack vector to jailbreaking and is usually manifested in the form of inputs sent to the AI system during use.

Jailbreaking can result in IP compromise and the failure of IP protections, exfiltration of the AI model and noncompliant, unsafe, or unethical use.

## 10.4 Securing AI Systems

There are unique challenges to securing AI systems that may not be addressed effectively by existing information security measures.

The field of Explainable AI seeks to give humans intellectual insight into AI systems through understanding the reasoning behind the output of AI systems. Most AI systems are not explainable and that poses at least two security challenges (and this list is not intended to be complete and undoubtedly there are threats yet to be discovered).

The first challenge is that if you cannot understand how a system works, then how do you know when its operation has been compromised? The outputs of AI systems are usually non-deterministic making it intuitively difficult to distinguish correct but unexpected output from the output of a compromised AI system.

The second is that if you do not know when a system was compromised, a consequence of not being able to know if it has been compromised, how do you restore it to pre-compromise operation? Standard disaster recovery strategies are predicated on being able to identify a pre-disaster system state.

For example, a ransomware attack can be mitigated by restoring a pre-compromise image of the data, an approach that is predicated on identifying an uncompromised image. But this approach will have mixed success with a poisoned AI system, given that the difficulty of detecting a compromise makes it hard to determine the pre-compromise state, and because the poisoning of the AI system may have occurred long before it manifests itself in a manner that is detected.

New security tools are needed that can prevent, detect, and remediate compromises of AI systems.

© SMPTE 2023, 2025 – All Rights Reserved

## 10.5 Agent Security

The two primary threats associated with agents are:

1. Malicious actors manipulating agents to perform in an unauthorized manner such as causing data exfiltration or disrupting a workflow.

2. Agents may perform unintended actions without the intervention of malicious actors, and these actions may threaten the security of workflows and assets.

The first threat is mitigated by security measures that protect the agents, the second threat is mitigated by security measures that protect the workflows and assets. However, observing that an agent's actions are not as intended, the non-deterministic nature of AI makes it difficult to tell these two threats apart.

Groups such as the Coalition for Secure AI are working to address these issues and many other AI security issues. (See https://www.coalitionforsecureai.org/leadership/).

AI security is made more difficult by the rapid development of AI capabilities. The ease of creating agents by developers not familiar with security and who are often using AI tools to write the code, amplifies the security threats.

Security by design principles has never been more important.

## 10.6 Security of Production System with AI Components

### 10.6.1 Zero Trust Architecture

The distributed nature of media production makes it complex to secure, and introducing AI systems into workflows adds to the complexity. MovieLabs' Common Security Architecture for Production[44] is a zero trust security architecture developed by MovieLabs and its member studios in partnership with cloud services providers for the purpose of securing existing and new workflows, whether running on a public or private cloud or in a traditional data center. At its core, zero trust requires universal authentication and authorization controlled by authorization policies[45] (called dynamic security policies in NIST SP 800-207[46]). An authorization policy sets out the conditions (such as who, what, when, and where) under which an action is authorized.

[44] Common Security Architecture for Production (CSAP) v1.3, 2023, https://movielabs.com/production-technology/production-security/

[45] CSAP uses the term "authorization policy" because the authors sought to avoid any ambiguity with the use of "security policy" in enterprise security rules and processes such as, "log off of your computer before going home".

[46] Rose, S., Borchert, O., Mitchell, S., Connelly, S. Zero Trust Architecture, NIST Special Publication 800-207. 2020. https://csrc.nist.gov/pubs/sp/800/207/final


© Copyright SMPTE 2023, 2025 – All Rights Reserved

### 10.6.2 Identity and Trustworthiness

An organization will evaluate whether a user is trustworthy before setting up an account for them in the identity management system. The identity management system's role is to authenticate that an entity requesting access to a system is the trusted entity they claim to be. How can we map that to AI? Before we can come to conclusions about the trustworthiness of AI systems, an existential question must be answered first: What is the thing, and what is the boundary of the thing, for which we wish to establish trustworthiness? Thus, the first problem is to define what identity of an AI system means.

Determining if an AI system, for example an AI agent, can be trusted[47] means knowing if, or the extent to which, an AI system that has not been compromised can be trusted to behave in an acceptable manner.

### 10.6.3 Authentication

In zero trust, authentication is required of users, systems, services and, potentially applications, which means that when it comes to AI, we need to be able to authenticate an AI participant (e.g., an AI system) whether it is acting autonomously or acting on behalf of another participant.

### 10.6.4 Authorization

Zero trust is deny-by-default which is a best practice in any security model, and authorization by an authorization policy is required before any action can be carried out. This intersects with the use of AI systems in two different ways.

The first is authorizing the actions of AI systems acting either autonomously as a participant or acting on behalf of another participant.

The second is when security is workflow driven, meaning that the authorization policies are generated at the behest of whatever is managing the workflow, which is an optimum way of enforcing the principle of least privilege[48] at the appropriate granularity. If the workflow is managed by AI systems, then authorization policies would be generated by AI.

[47] Trust is not binary nor is authentication when trust inference is used. A user's access might be limited to low value assets if the trust inference assesses there is a non-zero risk that the user isn't the trusted user it claims to be.

[48] A security principle that a system should restrict the access privileges of users (or processes acting on behalf of users) to the minimum necessary to accomplish assigned tasks. (taken from NIST SP 800-12 Rev 1).

© SMPTE 2023, 2025 – All Rights Reserved

# 11 The Impact of AI on the Media Industry

AI has become a crucial element in the media industry, revolutionizing a multitude of sectors including content production, audience analytics, content recommendation, and archives analysis. Figure 10 illustrates the interactions to deliver relevant content to the audience via linear or non-linear channels. This section provides an overview of the impact of AI on various facets of media and broadcasting[49].

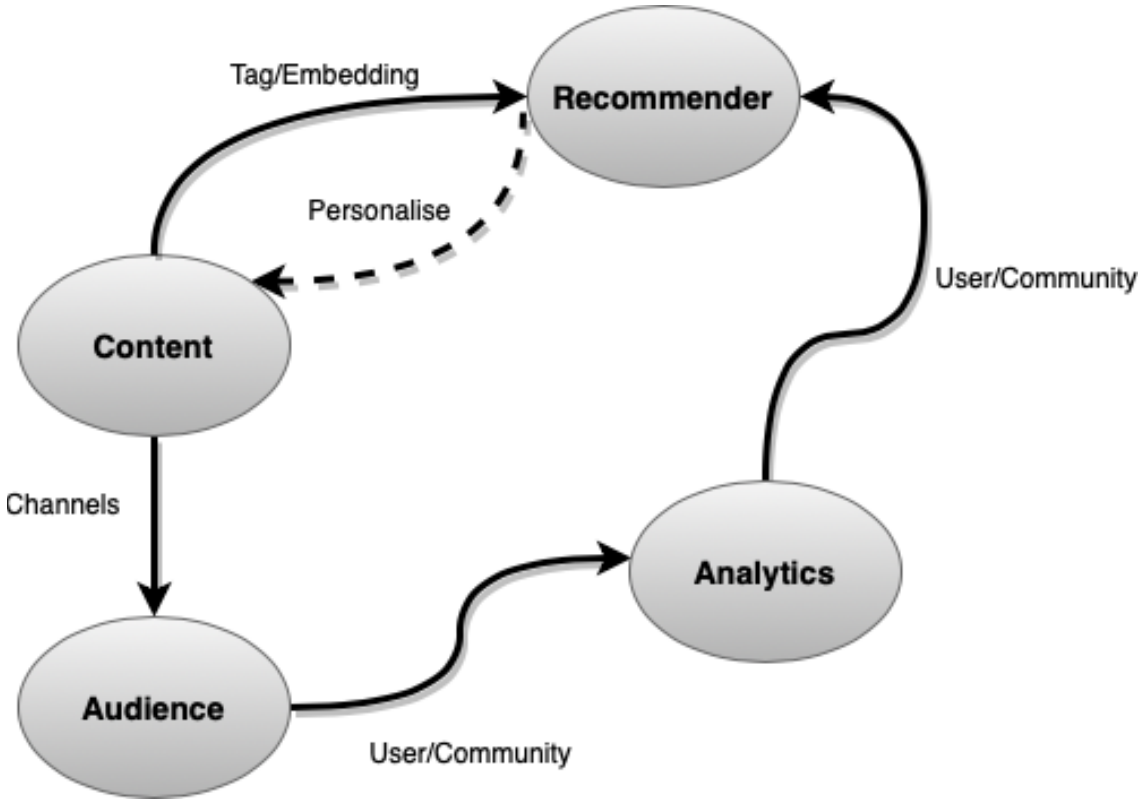


**Figure 10 — Closing the loop to reach audiences[49]**

[49] Rouxel, A., 2020, October. AI in the Media Spotlight. In *Proceedings of the 2nd International Workshop on AI for Smart TV Content Production, Access and Delivery* (pp. 1-2).


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 11.1 Content Production and Creation

In the realm of content production, AI enables automated production and content enrichment. AI's ability to automate production processes increases efficiency, making it possible to generate new types of content more quickly. Figure 11 shows an example of the capabilities of AI's real-time data display.


*Image: Second Spectrum*

**Figure 11 — Real-time statistics and content enrichment**

In sports video production, AI algorithms are used to produce near real-time content from 360° or autonomous cameras. To do so, the AI models must be able to recognize in-game situations and match highlights to produce relevant and high-quality content. Automated video editing uses similar technologies such as player tracking and highlight detection, paving the way for low-cost content publishing on social networks. For low-attendance soccer matches, automated content production and publication are already being used by broadcasters on their paid platforms. To take real-time content production a step further, AI is used to generate real-time statistics such as players' top speed, shot speed and passing probabilities to analyze the game and improve the user experience.

Furthermore, generative AI is revolutionizing content creation by automatically generating content, assisting in scriptwriting, fiction writing, music, images, and videos. AI's role in video editing is equally transformative, as it can analyze footage to produce coherent editing, increase productivity, and generate summarization for trailers, thumbnails, and sports event highlights.

© SMPTE 2023, 2025 – All Rights Reserved

## 11.2 Content Summarization, Metadata, and Annotation

AI plays a significant role in content tagging. First, the exponential growth of produced content requires automation of content tagging to enhance the content and make it searchable in archives to enable its repurposing. In the field of metadata, automatic metadata extraction is an important application that uses AI, utilizing facial recognition, speaker identification, landscape detection, object recognition, text tagging, and topic extraction from text. Generative AI may propose new ways to search for content based on multimodal description and similarity search, rather than explicitly tagging content. OpenAI CLIP[50] (Contrastive Language-Image Pretraining) and Google LiT[51] (Locked-Image Tuning) are multimodal models that perform tasks that require an understanding of both images and text. Such approaches can improve content retrieval and tagging for archives.

Video summarization is another process for adding value to content. It consists of slicing content by selecting sections that contain key points. AI can be used to suggest summaries to the editor or to generate them automatically. This technology allows content to be repurposed or trailers to be generated in different lengths. This is the key technology for repurposing content and automatically adapting content for posting on social networks. In the same category, automatic thumbnail extraction is widely used by broadcasters.

## 11.3 Audience Reach

AI helps to ensure that content reaches the right audience, at the right time, and on the right platform. This includes strategies for disseminating, republishing, or repurposing content.

Media companies are increasingly using AI to target audiences and optimize their content distribution strategy. Machine learning models analyze audience data to create audience segments and profiles, which are used to distribute content at the optimal time through the right channels and platforms. Audience segments form the basis of the repurposing strategy (i.e., republishing and refactoring content to target specific audiences on specific social networks, such as TikTok or Instagram).

AI-powered tools that combine ML and big data analytics enable real-time monitoring of social network trends, popular topics, and successful articles. Semi-automated in nature, these tools inform editorial decisions to identify the content that matches the trends. AI's pattern recognition capabilities enable media companies to correlate content with social trends. By using AI models such as predictive analytics, media companies can determine what content will resonate with audiences based on trending topics.

In summary, AI technologies such as ML, NLP, and predictive analytics play an essential role in improving content strategies and ensuring that media companies remain agile and responsive to audience behavior and preferences.

[50] https://openai.com/research/clip

[51] https://google-research.github.io/vision_transformer/lit/


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 11.4 Content Recommendation and Personalization

### 11.4.1 Overview

AI algorithms can identify patterns in user behavior to provide personalized content recommendations, enhancing user engagement and satisfaction.

Content recommendation and personalization play pivotal roles in the current age of information. As information and available content surge at an exponential rate, it becomes increasingly difficult for users to navigate this vast ocean of data. Content recommendation systems offer a solution to this overwhelming problem by providing more personalized and tailored content to each user

### 11.4.2 User Profile

A prerequisite for the process of content recommendation and personalization is understanding the user. The granularity of these analytics ranges from individual users to broader communities. User analytics focus on individual user behavior. It records a user's browsing history, time spent on pages, clicks, and other online activities. This data is a goldmine, as it provides the AI system with raw material to learn and predict a user's preference for future content. Community analytics go a step further, examining the behavior of a group of users who share similar interests. Studying communities can aid in understanding broader trends and patterns that may not be immediately apparent at the individual user level. They also help in providing recommendations to a user based on the actions of their community, even if the users themselves have not shown explicit interest in such content.

### 11.4.3 AI in Recommendation Systems

Recommendation systems are prominent applications of AI that rely on two main techniques: collaborative filtering and content-based filtering.

Collaborative filtering is a model built on the similarity between users, operating on the assumption that people who agreed in the past will agree in the future. This method is extremely potent, as it does not need to understand the content of the recommended object, making it versatile for various domains such as movies, music, and news.

Content-based filtering methods base their recommendations on the description of the items, pushing forward items similar to those that a user liked in the past. The description of the content, thus, becomes a key factor. Many systems now use AI-derived embeddings to recommend similar content. Embeddings are used to convert the description of items into a mathematical representation that machines can understand and compare.

One of the main pitfalls of recommender systems is the "bubble effect": the AI system adapts so well to user preferences that it tends to recommend only very similar types of content. This bubble can prevent users from discovering new and diverse content. In practice, however, recommendation systems can combat the bubble effect by incorporating strategies such as diversity and randomness to ensure that content is not only relevant, but also diverse and sometimes surprisingly well matched. In addition, the use of hybrid filtering techniques that combine collaborative filtering and content-based filtering, and the provision of user-controlled parameters, further increases the diversity of recommendations. This diversity of content is achieved by providing a mix of popular trends and personal tastes, while allowing users to adjust the degree of novelty of their content.

© SMPTE 2023, 2025 – All Rights Reserved

### 11.4.4 Context-Aware Recommender Systems

Context-Aware Recommender Systems[52] (CARS) extend traditional recommendation systems by considering both the user's past behavior and the current context. This contextual data can include a wide range of factors such as the user's current location, time of day, prevailing weather conditions, and significant recent events. For example, a video recommendation system, such as those on streaming platforms, could adjust its recommendations based on the day of the week and time of day. Additionally, during a major sporting event such as the World Cup or Super Bowl, the system could prioritize related content, recaps, or related documentaries to capitalize on the heightened interest in these events.

In essence, the integration of AI into content recommendation and personalization is reshaping the way users interact with content. It touches many aspects of the user profiles and the content consumed, influencing users and society at a larger scale.

Standardization opportunities for recommender systems are covered later in this report.

# 12 AI Ethics

## 12.1 What is AI Ethics?

AI ethics is the set of ethical considerations involved in the design, development, and deployment of artificial intelligence systems. Because it is hand-built, by humans for humans, AI architectures encode organizational values and human biases at all levels, from data collection to model deployment.

As such, AI ethics is concerned with virtually all aspects of development: design for social and environmental good, respect for privacy and minority representation in data collection, racial, gender, and cultural bias in training data, inclusivity in data science teams, biases in machine models, the need for transparency and explainability, and of course, benevolence of end-uses. AI ethics is a vast ecosystem of practices, systems, and goals.

In the media industry, this means that ethical considerations are present throughout the data science pipeline. Especially where data informs decisions. Some ethical considerations are more strategic, such as hiring diverse teams or setting up robust processes to protect consumer privacy. Others, such as reviewing minority representation in training data, are more tactical and should be embedded in data science and product development teams. AI ethics is an emergent property of the machine intelligence pipeline, arising from ethical considerations at all levels of the organization.

As data science spreads throughout media-making decisions and workflows, it is important for organizations to develop the confidence that their data and models are fully understood. This means they are transparent and auditable, trusted, rid of unseen biases, and do not result in privacy violations or discriminatory outcomes.

Direct-to-consumer business models and the considerable opportunity presented by the emergence of multiverse environments are pushing the industry much deeper into the arms of machine models, thus creating even stronger ethical requirements. Computer vision models, large language models, and generative models, have all crossed a threshold of performance that carries opportunity and ethical risk. "Deepfakes" arise from a powerful but ethically dangerous technology. Synthetic characters and agents (chatbots) pose considerable ethical dilemmas.

[52] Adomavicius, G. and Tuzhilin, A., 2010. Context-aware recommender systems. In *Recommender systems handbook* (pp. 217-253). Boston, MA: Springer US.

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

## 12.2 Why Should Ethics Be Useful in AI Development?

### 12.2.1 Because AI Is Early, Critical, and Misunderstood

AI is at the same time disruptive, vague, complex, experimental—and a great story. It is difficult to understand, and easy to load up with fears and fantasies.

This is a dangerous combination. The convergence of corporate hype, fledgling methods, biased datasets, and the urgency to productize, are all fertile grounds for failure—and failure is generally good with tentative tech like AI—except when models are put in a position to make decisions about policing, hiring, synthetic conversations, or even content recommendation and personalization. Then, failure may come at a high human cost.

The time to discuss ethical considerations in AI is now, while the field is still nascent, teams are being built, products roadmapped, and decisions finalized.

### 12.2.2 Because It's the Law

According to the United Nations Conference on Trade and Development, 77% of all UN member states already have data privacy laws or have pending legislation. GDPR (European Union), and the joint CCPA/CPRA in California have already alerted the private sector about the attention regulators are paying to consumer data and artificial intelligence-driven decisions. But it seems the trend is growing and spreading around the world. The City of Los Angeles recently took legal action against IBM for misappropriation of user data for the latter's weather app. Goldman Sachs has been investigated for discrimination against women in some credit card applications. The list goes on, and it will get much longer.

### 12.2.3 Because Failing Is Expensive

AI development is no longer just a technical issue; it is increasingly becoming a risk factor. Because AI is experimental, impactful, and expensive, organizations must examine the downside risk of deploying underperforming and unethical AI systems, especially because, in most cases, ethical and technical requirements are the same. For example, unseen bias is as bad for model performance as it is discriminatory. Model transparency is not just an ethical consideration—it is a trust-building instrument.

To help organizations identify, assess, and mitigate AI-related risks, the MIT AI Risk Repository[53] provides structured insights through three key components: the AI Risk Database, the Causal Taxonomy, and the Domain Taxonomy. The AI Risk Database systematically links each identified risk to supporting evidence and categorizes them according to their nature and context. In addition, the repository's Causal Taxonomy classifies risks based on how, when, and why they occur, while the Domain Taxonomy organizes risks into specific domains and subdomains for targeted analysis.

In 2017, Amazon had to famously scrap costly machine-driven job applicant processing software because it discriminated against women, as it was trained on an overwhelmingly male dataset. This cost the company in three ways: (1) the obvious reputation hit for a computing leader, (2) the cost of developing, then scrapping, a faulty application, but most importantly, (3) the opportunity cost of making bad decisions based on biased machine learning models (i.e., trained on statistically biased data). The following year, an autonomous vehicle tested by Uber killed a pedestrian in Arizona, in part because its model had not been properly trained on jaywalking samples.

[53] See https://airisk.mit.edu/

© SMPTE 2023, 2025 – All Rights Reserved

### 12.2.4 Because Media Needs Its Own Voice

The media and entertainment industry is equal parts a creative industry and a tech industry. As such, it has its own voice, its own culture, and nearly 150 years of success marrying human and technological genius. It also holds a substantial and powerful place in our society as the mass distributor of human narratives and social norms.

Media must bring this unique voice and hybrid human/machine culture both to AI development and the debate on AI ethics. And as the industry starts developing and deploying AI applications from development to distribution, there is a need to approach this issue at the industry level first.

Media and entertainment companies collect and process large amounts of consumer data, for example. Increasingly, this means that they must comply with a growing list of legal regimes and data governance requirements. Similarly, there's a substantial opportunity to use computer vision in the production (virtual production) and post-production processes (color correction, translation and localization, and of course, VFX work).

The quality and diversity of training sets, how color correction can affect representation of minorities, and of course the use of "deepfake" technology, are all critical areas where ethical considerations are paramount. The media industry's history of sophisticated legal practice around likeness rights, royalties, residuals, and participations, is a substantial advantage in navigating issues related to computational derivatives of image and content.

A standards-based approach to verification and identification is key, and not only of the image (e.g., format and technical metadata), but also of the talent itself and the attestations of authenticity (i.e., quantitative notability, via citations and historical credits information). Persistent, interoperable, and unique identifiers have aided media supply-chains in the past, and could well help with the labeling and automating the provenance of authentic talent in the future age of AI in media and entertainment.

At a minimum, requirements for data and model transparency would go a long way towards reinforcing trust in computational methods and help convert those in the industry still reluctant to use statistical learning to optimize human processes. Around the corner, the development of conversational agents (chatbots) creates serious ethical risks, especially as the industry looks to create highly immersive and personalized experiences in the multiverse.

### 12.2.5 Because It's Fundamental

As mentioned, technical and ethical standards in AI are overwhelmingly one and the same. Bias is the model-killer. Black box algorithms help no one. Intellectual and cultural diversity is critical to high performance. Product teams must broaden their ecosystem view.

Entangling the ethical implications of AI with product goals goes a long way towards deepening our collective understanding of the field. And thinking about AI ethics forces us into systems thinking, which is an almost Darwinian imperative in all areas of contemporary business, technology, and society.


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 12.3 Some Core Principles

In his excellent book, "*Trustworthy Machine Learning*,"[54] IBM researcher Kush Varshney compares the challenges of trustworthiness in AI and machine learning to those of processed foods. In the early 20th century, processed food companies like Heinz had to gain the trust of consumers and regulators through "unadulterated ingredients, transparent containers, sanitary food preparation, factory tours, labels, and tamper-resistant packaging." Inspired by this effort to build trust in an essential component of society (food), following are some core principles of AI ethics for the media industry.

### 12.3.1 Broadness

AI ethics should be viewed in the larger context of computational ethics. Whether or not they are built upon AI or ML architectures (increasingly they are), systems with impact on an organization, a business model, a revenue stream, a key life decision (such as hiring or getting a mortgage), a minority group, or medical treatment, are subject to bias. As such, they need to be understood, built transparently, and audited regularly. Computational bias, AI or not, means bias on a massive scale. Ethical considerations should be a systematic part of all aspects of digital product design, development, and QA. This seeding of ethics at the product level is essential to view bias as a complex ecosystem of inputs, features, models, outputs, ... and outcomes.

### 12.3.2 Fit

We are what we build. Any organization’s output, products, and decisions (deliberate or not) inherently fit its culture and values. This is why AI ethics is high-stakes: it deploys an organization's culture and values on a large scale. Because they shape society at scale and have a history of taking the public interest seriously, media companies have a distinct responsibility to move forward with their AI ambitions, with full awareness of these applications' ethical considerations. They should ensure that all aspects of their development (including data collection), deployment, and end-uses, support the law as well as their own values regarding privacy, justice, tolerance, and human rights.

### 12.3.3 Inclusivity

Gender, racial, social, intellectual, and cultural diversity of all kinds are critical to maintaining a richness of voices, societal experiences, and cultures when developing AI systems. It's not just the right thing to do, it's the smart thing to do. AI and machine learning are extremely difficult to get right. They touch a large number of technical, academic, cultural and intellectual domains. The more diverse the voices, the more viewpoints, the more creative solutions, the more chances of success. Diversity creates richness in products and organizations, and is a critical factor in the performance of data science teams. It is also a bulwark against confirmation bias, which can be costly once enshrined in organizational processes and systems.

### 12.3.4 Transparency and Trust

The entire value chain of AI development, from product design to data collection to model deployment, should be secure, transparent, explainable, and auditable. Black box machine learning frameworks are both ethically and statistically dicey. They foster sloppiness in data science teams and mistrust for those already suspicious of machine models. What cannot be explained should not be deployed in a decision-making environment.

In a world where organizations are often too suspicious or too enthusiastic, only secure, transparent, explainable, and auditable machine models can scale resiliently. Additionally, all stakeholders deserve transparency, each in their own language, across different points of view and technical sophistication. Ethics should be part of Quality Assurance for any and all computational systems.

---

[54] Available at http://www.trustworthymachinelearning.com

© SMPTE 2023, 2025 – All Rights Reserved

### 12.3.5 Openness

AI is still a technical, ungoverned frontier. Everything around it, from roadmapping to modeling to seeding in company culture, is complex and challenging. Mistakes will happen. Organizations must communicate comprehensively and with humility about their journey to approach and implement processes around ethical AI, for the benefit of all.

After all, we are all trying to make ethical something that even experts still cannot fully understand. Transparency will help regulators, senior business executives, and the general public understand that artificial intelligence is the exact opposite of “magic”. It’s either blood, sweat, and tears … or it’s not AI.

Just as with technical and organizational implementation, ethical considerations in the development and deployment of AI are complex and laborious. Being open and didactic will not only feed the public debate about AI with realistic and trustworthy narratives (as opposed to noxious hype), but will create a collective mindshare for organizations to learn from each other’s successes and failures.

## 12.4 The AI Ethics Pipeline

Most of the challenge lies with implementing the previous principles. AI is an emerging technology, and AI ethics is an almost entirely blank slate. Examples of successful, organization-wide implementation of machine learning transparency and trustworthiness are extremely rare. Nonetheless, some early experimentation in the pharmaceutical and financial services industries have generated some best practices.[55, 56]

### 12.4.1 Organization

This is perhaps the most critical step in laying out an AI ethics strategy, because nothing is more impactful—and difficult to build—than organizational systems, incentives, and mechanisms. This is where AI ethics becomes enshrined. Following are a few principles borrowed from successful experiments deploying AI governance in a corporate environment:

**A. Set clear goals but flexible roadmaps.** AI ethics is a nascent and uncertain practice that touches upon virtually every business process. It needs flexibility to experiment and diverse buy-in to flourish. It is a good idea to have open and transparent conversations at all levels about expectations prior to setting a roadmap. In media, it means that virtually all sectors across marketing, development, and technical implementation have a piece of the puzzle and a role to play in setting expectations for an AI ethics initiative. Also, the AI ethics work is never done, it will be a perennial trial and error process.

**B. Inventory organizational resources already available to seed an AI ethics program.** Chances are that a foundation of an ethics practice already exists within the organization. Set up an executive committee inclusive of all voices, business units, technical backgrounds, and cultures. Promote the initiative (and the group) internally. Educate and train to create a level playing field within the organization. Set up an executive committee inclusive of all voices, business units, technical backgrounds, and cultures. Promote the initiative (and the group) internally. Educate and train to create a level playing field.

**C. Create clear lines of responsibility and accountability.** Ethics is funded, incentivized and supervised at the corporate level, but its implementation must be bespoke to the needs, resources and priorities of each business unit. Product managers in each business unit should be front and center in leading the deployment of ethics policies and practices.

---

[55] https://www.riskinfo.ai/post/developing-trust-in-ai-for-financial-services-current-progress-and-future-directions

[56] https://www.caro.vc/


© Copyright SMPTE 2023, 2025 – All Rights Reserved

**D. Foster cultural and intellectual variety.** Because of the multifaceted nature of AI ethics, working groups should include a wide variety of stakeholders. This obviously means age, gender, racial and cultural diversity, but not only. Consumer research teams, product managers, legal and compliance teams, and data scientists each bring a different perspective to balancing the requirements of ethics with that of performance and customer experience. Ethics should not be the exclusive domain of those preoccupied with governance and risk.

**E. Communicate with senior stakeholders.** Learn how to converse about AI and ethics with senior business executives. C-suite and legal executives seeking clear and measurable ROI in their AI efforts—including ethics—require informed guidance on risks, given the experimental nature of this technology.

**F. Make the process as transparent and measurable as possible.** Measurement is important in any trial and error process. So is transparency about mistakes and lessons learned with regards to building ethical and trustworthy AI applications. It is a completely new domain related to a completely new technology. Failure will happen.

### 12.4.2 Product Design

Safety and transparency need to be fundamental product goals for designers of AI systems. AI-driven guardian agents[57] have recently come into use for monitoring other AI. These agents are designed to track the behavior and performance of an AI system; evaluate quality, safety, and accuracy; and make appropriate interventions by blocking or redirecting harmful or inappropriate actions. By automating oversight of mundane tasks, guardian agents can be an effective first line of defense and help administrators focus on the bigger picture of making AI systems more trustworthy.

Whoever is given the lead to examine ethical considerations in computational systems should start by laying out requirements of "AI trustworthiness and transparency" that are specific to each stakeholder. A customer will have different needs, and speak a different language, than a marketing executive, or a data scientist. And ethical considerations even vary within customer types. For example, audiences must trust an AI-driven recommendation algorithm to "know" their specific tastes, while intelligently expanding their creative horizons.

Marketing executives and analysts must trust that a sentiment analysis engine correctly distinguishes positive from negative sentiment. Deeply semantic sentiment domains like sarcasm are still difficult to measure. A digital product manager must trust that their virtual character won't stray into inappropriate conversations with users.

Listing all stakeholders and analyzing their various cultures and needs are useful initial steps in the AI ethics pipeline. Performance and transparency are also major components of trustworthiness. It is critical to associate AI and ML models with their quantitative performance.

Because computational ethics (not just in AI) are implemented in tandem with those with responsibility over the use case behind AI applications, two roles within organizations take a pivotal role in AI ethics implementations: insights leaders and product managers.

The first one (internal facing, focused on processes) is straightforward; it is the responsibility of the head of insights to ensure that insights are collected, processed, and output transparently and ethically. Data and model integrity are core responsibilities.

[57] https://www.gartner.com/en/articles/guardian-agents

© SMPTE 2023, 2025 – All Rights Reserved

Product managers could also take a central role in leading external-facing (product-focused) AI ethics considerations. Because they are by nature systems thinkers, care about the customers at least as much (if not more) as about the company, and are ultimately accountable for an organization's raison d'être (its products), product leaders are essential "quarterbacks" of computational ethics. They are best positioned to weigh all considerations of transparency, integrity, and functionality. Plainly put, they sit at the intersection of product and users, and can best weigh user experience vs. ethical requirements.

Identify clearly where technical, business and ethical goals are aligned, and where they are not. Start with the former, and promote quick wins with low hanging fruit (see the hierarchy of needs pyramid[58]). Model performance and sample bias are examples of ethical issues where technical, business, and ethical goals are aligned. Other issues, such as the decision of whether or not to use "deepfake" technology in the VFX process, may require more extensive consultations with C-level executives. While some products may perform better with more intrusive consumer data collection, this could be seen as a user experience problem. Users may have the choice between an enhanced experience that collects data more aggressively and a limited experience that protects privacy.

Finally, an effective AI ethics program will make checking for bias, trustworthiness, and transparency a function of quality assurance. This is a critical part of the ethics pipeline, as it ensures balance between those requirements and the needs of user experience and product performance.

### 12.4.3 Data Collection

The data collection process is very much at the heart of the AI practice as a whole, and of ethical considerations in particular. "Garbage in, garbage out" is the Golden Rule of data science: models are only as good as the data on which they are trained. Identifying biases in data collection and monitoring how bias might increase over time are at the heart of the AI ethics practice.

**1. Know your data.**

This is perhaps the most important part of the AI ethics pipeline. It is also a major area where statistical and ethical requirements are one and the same.

Data is the raw material of data science, and it is data scientists' first and foremost responsibility to know their dataset, its strengths and weaknesses, inside and out, to be able to map issues with a skewed output (the model) back to skewed inputs (the data). Using representative datasets that fully take into account gender, race, culture, etc., is not just the right thing to do, it is the statistically sound thing to do.

Sample bias is a primary source of poor ethical outcomes in AI. For example, facial recognition applications have been notoriously underperforming in the detection of both darker skin tones and females (see Joy Buolamwini's and Timnit Gebru's "Gender Shades" study), due to substantial under-representation of darker-skinned samples in computer vision training sets.

In their 2018 paper, Gebru and Buolamwini noticed that two of the most prominent training sets of faces at the time, IJB-A and Adience, are composed of 79.6% (IJB-A) and 86.2% (Adience) of lighter skinned faces. The maximum error rate for lighter-skinned males in these models was 0.8%, vs. 34.7% for darker-skinned females.[59]

[58] https://shopify.engineering/shopify-unique-data-science-hierarchy-of-needs

[59] http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf


© Copyright SMPTE 2023, 2025 – All Rights Reserved

Similarly, the bias in Amazon's hiring software came from the fact that it had been trained on resumes received by the company over a 10-year period. An overwhelming majority of these resumes were from men. As a result, it penalized resumes that included the word "women's", as in "women's basketball team".

This is a complex and multi-layered issue to tackle. For example, even subtle nuances in how data is collected (the way a question is asked, or how incentives to contribute one's data are structured) can dramatically affect the resulting dataset. Knowing how the data has been collected, and what biases may lie within that process, is increasingly a core responsibility of data scientists.

**2. Know your problem.**

AI and machine learning are used to solve real-world problems by using data to represent and model those problems in their larger context. This is identical to how the human brain functions; we use data to model (summarize) an infinitely complex world, and use those models to act upon the world. Knowing intimately the problems, systems, behavior, phenomena they are trying to model, and in this case, what biases may be inherent in them, is a key strength of great data scientists. Some parts of our world are simple, but most are extremely complex, not least of which human behavior. When modeling a real-world system (such as audience decisions), the data scientists need to understand that the set of variables they are analyzing accurately represents the larger system about which they are trying to generalize their findings. They also need to identify if and when the inequities and/or biases of the system itself will be passed on to the model. Oftentimes, the set of variables available to the system is too small and partial for the model to generalize to a much more complex (and quickly evolving) real-world system. Over- and under-fitting are the statistical manifestations of this issue, as is encoding real-world biases in machine models.

This is especially important when modeling human behavior, specifically during audience research. Posting a thread on Reddit or reposting a post on X or liking a post on Facebook are three radically different types of social behavior expressed by different genders, age groups, races, and subcultures. And models about them generalize differently, which is why it is critical, in the practice of AI ethics, to maintain intimate knowledge of how underlying social, cultural or behavioral biases may impact data collection. It is critical for data scientists to understand the underlying biases not only in the input data but in the systems they are trying to model.

This is best done as a collective process, since confirmation biases (the tendency to look for, cherry pick, or interpret insights according to one's preexisting beliefs) are also present in data science teams, or quantitatively-driven functions such as marketing and consumer research.

**3. Communicate clearly about use (opt-in), biases, and their tradeoffs.**

When collecting user data, opt-ins are a must; they are also increasingly required by law. The best practitioners in this domain avoid legal language and use instead a simple user-facing explanation of how personal data will be used, and what the implications of opting in and out are for the user experience (for example, how opt-out of sharing data would affect personalization).

Bias often cannot be avoided, in which case it is critical for data science teams to communicate fully and clearly to end-users about how bias impacts the skew of their model. A simple annotation in the output can be very powerful in building transparency and trust, without which no culture of data (let alone AI) can be successfully scaled in media organizations.

© SMPTE 2023, 2025 – All Rights Reserved

### 12.4.4 Modeling

In ethically compliant AI or machine learning, building trust is both critical and labor-intensive. There are two parts to this: creating explainable statistical models (model transparency), and effectively reporting key features of the model, so it can be quickly audited by end-users for bias and potential performance variations.

**1. Model Transparency**

Transparency means building simple models to explain complex ones, to give data scientists a window into often complex and interlocking machine learning architectures. This is increasingly a challenge, as neural net architectures become deeper and more integrated with other types of models. Luckily, the past few years have seen a flurry of development of AI transparency tools. All providers of cloud-based machine learning have started offering tools to interpret and understand their models (for example, AWS's Sagemaker model explainability feature - based on SHAP, Microsoft's InterpretML toolkit, or Google AutoML's feature scoring tool). These are useful because variables in a model are hierarchical (some are more powerful than others within the model), and surfacing that structure is a key step in understanding how the power of certain variables related to gender, race, or culture, for example, may perpetuate inequalities.

The following are some popular explainable AI tools:

- **SHAP (SHapley Additive exPlanations):** Perhaps one of the most widely known AI transparency tools, because it works across a wide range of models, from linear regression to deep learning, and covers a wide variety of domains, including computer vision and NLP. SHAP uses a game theoretic approach to rank features (for example, words in a sentiment analysis model) by order of importance in predicting the output. This is the kind of hierarchical view that is very helpful not just in the context of AI ethics, but for data scientists to QA their own models. Transparency is one of many areas serving both ethics and model performance.

- **LIME (Local Interpretable Model-Agnostic Explanations):** Similar to SHAP, but more computationally efficient (faster). LIME also ranks features by how much it contributes to the output. For example, in an image classifier, it can produce a heat map of an image with "useful" features in green and "not useful" features in red. SHAP can do this as well. LIME is very popular for Python's scikit-learn users because of its built-in integrations.

- **ELI5:** Works similarly to SHAP and LIME, but is perhaps one of the most popular transparency packages in Python, because of its integrations across the board with scikit-learn, XGBoost, Keras, and others.

- **AIX360:** Developed by IBM Research but still open source, this toolkit has extensive functionality and is not dissimilar from Google's "what if" tool, but can be used outside of the Google CloudAI environment, although it is not for beginners.

- **Google's "What-if Tool":** Allows data scientists to test a model's performance under a variety of different situations. It helps to understand the impact of various variables (such as race or gender) on the model itself. This is an excellent and intuitive tool for beginners using the Google Cloud AI infrastructure, as it has a user-friendly visual interface and can be run easily (and with minimal code) from platforms such as Jupyter Notebooks, Google Colab, and even Tensor Flow's TensorBoard dashboard. It can be used at various stages of the data science workflow, can support TensorFlow models out of the box, and works with tabular, image, and text data.

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

Transparency is not just key; it is a perennial concern. The world changes, the problem changes, the data changes, and model performance is affected. There is no longer a fit between the model and the system, or behavior that it represents. "Model drift", as it is referred to in data science circles, impacts ethical outcomes, because what may be ethical in January may no longer be in June. Only transparent and auditable models can catch model drift before it causes damage.

**2. Model reporting: model cards**

Too often, machine models are released with incomplete documentation and unclear context. As a result, they are applied to contexts in which they do not perform well or are not appropriate in which to deploy. Created by Margaret Mitchell and Timnit Gebru's team at Google in 2018, "model cards" are standardized documentation laying out all of the information necessary to evaluate a model and benchmark its performance in a variety of contexts. To be sure, libraries and models often come with documentation, but it is often incomplete, too long, and generally could benefit from standardization. Model cards are standardized "food labels" for data science that also—ideally—benchmark a model's performance in a variety of contexts and use cases, some related to inclusion and bias.

Per the Mitchell/Gebru/team paper ("Model Cards for Model Reporting"(https://arxiv.org/pdf/1810.03993.pdf): "Model cards are short documents accompanying trained machine learning models that provide benchmarked evaluation in a variety of conditions, such as across different cultural, demographic, or phenotypic groups (e.g., race, geographic location, sex, Fitzpatrick skin type, and intersectional groups) that are relevant to the intended application domains. Model cards also disclose the context in which models are intended to be used, details of the performance evaluation procedures, and other relevant information."

# 13 AI standards landscape

## 13.1 State of Play

AI is already widely used today. However, technical standardization of AI is still in its infancy. AI standards should provide an agreed-upon language and frameworks to support the development and deployment of technological innovations. In Europe, ETSI and CENELEC have published ambitious standardization programs, partly stimulated by the standardization framework proposed by the EU AI Regulation. ETSI focused on safety issues related to AI and machine learning, while CENELEC focused on trustworthiness and ethics. More globally, ISO/IEC JTC 1/SC 42 Artificial Intelligence is charged with driving JTC1's artificial intelligence standardization program and to provide guidance on the subject to IEC and ISO committees developing artificial intelligence applications[60]. More specific to media and entertainment, ISO/TC 36 Cinematography has established a liaison relationship with SC 42 (as has SMPTE) and formed an AI study group to assess the AI landscape for cinema and recommend a work program if warranted[61]. In addition to security, trust and ethics, the standardization of AI is likely to have a significant impact on the cross-sectoral use of AI. Many solutions currently used in AI are single-sector solutions, for example, for healthcare.

[60] https://www.iso.org/committee/6794475.html

[61] https://www.iso.org/committee/48090.html

© SMPTE 2023, 2025 – All Rights Reserved

AI technical standards cover a wide variety of considerations, including data management, system accuracy, operating range of the AI systems, interoperability, safety, and reliability. Technical standards can provide guidelines for the development, evaluation, and interoperability of AI systems, as well as a methodology for ensuring system quality, maintainability, and portability. Standards ensure the quality as well as the scope and limitations of systems and also provide a means of measuring that systems comply. For example, there may be a legal requirement for algorithmic transparency. However, without defining what algorithmic transparency is and how to measure it, it can be difficult to assess whether a particular AI system meets these requirements and compare it to another similar system. This difficulty can discourage adoption of the technology. For this reason, in many cases, technical standards are a key element in determining whether an AI system is appropriate for use in a particular context.

## 13.2 Regulation Background

In the United States, the National Institute of Standards and Technology (NIST) has developed the AI Risk Management Framework[62] (AI RMF), which takes a different approach to the EU AI Act, particularly in terms of risk management, stakeholder involvement, and common vocabulary. The AI RMF aims to manage risks from all AI systems, regardless of their level of risk. It provides a structure for identifying potential harms at all levels of society and engages multiple stakeholders in the process. The AI RMF is organized into four functions: map, measure, manage, and govern, each with detailed categories and subcategories. NIST has published a profile (NIST AI 600-1[63]) for Generative AI, which addresses risks specific to the use of LLMs, cloud-based services, and acquisition.

The EU AI law, however, focuses primarily on high-risk AI systems. The implementation of the AI RMF is incentive-driven, meaning that the self-interest of companies plays a significant role in the adoption of this framework, whereas the EU AI Act takes a more legislative approach and is standards-oriented[64].

## 13.3 AI Discussion Hubs

In addition to standards bodies, many organizations offer AI exchange platforms to track the evolution of AI and its applications and societal challenges, including the following:

- **AI for Good:** Aims to harness AI for social innovation, with a focus on the environment, health and education: https://aiforgood.itu.int
- **OECD:** Provides multidisciplinary data and analysis on artificial intelligence as well as a dialogue on AI. Provides a tool for tracking AI research activities: https://oecd.ai/
- **Partnership on AI (PAI):** A nonprofit partnership of academic, civil society, industry, and media organizations creating solutions to make AI work for people and society. The AI and Media Integrity Program develops best practices to ensure that AI positively impacts the global information ecosystem: https://partnershiponai.org
- **AI Watch:** Created by the EC, monitors the development, adoption and impact of artificial intelligence in Europe, publishing dozens of reports, including the AI Index: https://ai-watch.ec.europa.eu/index_en

---

[62] AI RMF: https://www.nist.gov/itl/ai-risk-management-framework

[63] NIST AI 600-1 Artificial Intelligence Risk Management Framework: Generative AI Profile, July 2024. https://doi.org/10.6028/NIST.AI.600-1

[64] Harmonising Artificial Intelligence: The Role of Standards in the EU AI Regulation. Mark McFadden, Kate Jones, Emily Taylor and Georgia Osborn Oxford Information Labs December 2021 https://oxcaigg.oii.ox.ac.uk/wp-content/uploads/sites/11/2021/12/Harmonising-AI-OXIL.pdf


© Copyright SMPTE 2023, 2025 – All Rights Reserved

- Open source initiatives (as discussed in Section 2.2) often serve as standards incubators because that ecosystem is not constrained by the due process rigors of de jure standardization. As a pre-standardization activity, open source projects can address interoperability and related “pain points” more rapidly by serving as de facto standards until formal standards are adopted.

## 13.4 Standardization Policy

- CEN/CENELEC Focus Group on AI, CEN-CLC/JTC 21, focuses on producing standardization deliverables that address European market and societal needs, as well as underpinning EU legislation, policies, principles, and values[65].

- The White House in the United States has released a plan[66] to prioritize development of domain-specific standards for artificial intelligence led by NIST.

- The European Commission (EC) is considering regulating AI using a risk-based approach. Members of the European Parliament believe the EU should act as a global standard-setter for AI. The EU should not only regulate AI as a technology, but the level of regulatory intervention should be proportionate to the type of risk associated with the particular use of an AI system. The European Commission has proposed a legal framework for artificial intelligence that aims to address the risks posed by specific uses of AI through a set of rules focused on respect for fundamental rights and safety. The EC intends to address the risks posed by certain uses of AI with a set of rules that will give Europe a leading role in setting the global gold standard. This framework gives AI developers, implementers and users the clarity they need by intervening only in cases not covered by existing national and European legislation. The AI legal framework provides a clear and easy-to-understand approach based on four different levels of risk: unacceptable risk, high risk, limited risk and minimal risk. This global policy will be accompanied by practical standards[67].

[65] https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/

[66] America’s AI Action Plan: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf

[67] A European Strategy for Artificial Intelligence: Lucilla SIOLI https://www.ceps.eu/wp-content/uploads/2021/04/AI-Presentation-CEPS-Webinar-L.-Sioli-23.4.21.pdf

© SMPTE 2023, 2025 – All Rights Reserved

## 13.5 Overview of AI Standards

In its report[68], EU Joint Research Centre provides an overview of standardization and classifies the standards in seven groups.

1. Data and data governance, related to:
    a. Data management practices;
    b. Dataset definition for training and testing;
    c. Data quality criteria.
2. Risk management system
    a. Identification of risks associated with AI system, adequate design and development, system compliance (with previous requirements), for instance, accessibility to children
3. Technical documentation and record-keeping
    a. General description of the AI system, detailed information about the monitoring, functioning and control of the AI system, and a list of the harmonised standards applied in full or in part
    b. Record-keeping
        i. Logs, dataset, identification of the natural persons involved in the verification for high-risk applications
4. Transparency and provision of information to users
    a. Instructions for use, (AI system) intended purpose, performance for targeted users, human oversight measures
5. Human oversight
    a. AI system's output, automation biases, capacities and limitations of the AI system, ability to stop the operation of the high-risk AI system
6. Accuracy, robustness, and cybersecurity
    a. Declaration in the instructions of use, system vulnerabilities exploitation, training datasets manipulations
7. Quality management system
    a. Policies, procedures, design, design control and design verification, data management (including: data collection, data analysis, data labelling, data storage, data filtration, data mining, data aggregation, data retention), accountability framework

[68] Nativi, S. and De Nigris, S., AI Standardisation Landscape: state of play and link to the EC proposal for an AI regulatory framework, EUR 30772 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-40325-8, doi:10.2760/376602, JRC125952


© Copyright SMPTE 2023, 2025 – All Rights Reserved

Table 2 provides a summary of the relevant standards for the AIA[69] key requirements.

**Table 2 — Relevant standards for the AIA key requirements [68]**

| Requirements: / Standard: | Data and data governance | Technical documentation | Record-keeping | Transparency and information to users | Human oversight | Accuracy, robustness, and cybersecurity | Risk management system | Quality management system |
|---|---|---|---|---|---|---|---|---|
| ISO/IEC TS 4213 | ✓ | | | | | ✓ | | |
| ISO/IEC 5259-2 | ✓ | | | | | | | |
| ISO/IEC 5259-3 | ✓ | | | | | | | ✓ |
| ISO/IEC 5259-4 | ✓ | | | | | | | ✓ |
| ISO/IEC 5338 | ✓ | | | | | ✓ | ✓ | ✓ |
| ISO/IEC 5469 | ✓ | | | | | ✓ | ✓ | |
| ISO/IEC 23894 | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ISO/IEC 24027 | ✓ | ✓ | | ✓ | | | | |
| ISO/IEC 24028 | | | | ✓ | | | | |
| ISO/IEC 24029-1 | ✓ | | | | | ✓ | | ✓ |
| ISO/IEC 24668 | ✓ | | | | | ✓ | | |
| ISO/IEC 38507 | ✓ | | | ✓ | ✓ | | ✓ | ✓ |
| ISO/IEC 42001 | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| ETSI SAI 002 | ✓ | | | | | ✓ | | |
| ETSI SAI 003 | | | | | | ✓ | | |
| ETSI SAI 005 | ✓ | | | | | ✓ | | |
| ETSI SAI 006 | | | | | | ✓ | | |

[69] AIA: Artificial Intelligence Act. In 2021, the European Commission issued the proposal for a Regulation laying down harmonised horizontal rules on artificial intelligence (i.e., the Artificial Intelligence Act: AIA). https://artificialintelligenceact.eu/ai-act-explorer/

© SMPTE 2023, 2025 – All Rights Reserved

## 13.6 Examples of AI Standards

**ISO/IEC TR 24029-1:202: Assessment of the robustness of neural networks**
Software validation is a critical part of releasing software into production. AI systems are also subject to validation; however, the current techniques used in AI systems pose new challenges that require specific approaches to ensure proper testing and validation. ISO/IEC TR 24029-1:2021 provides general information on existing methods for assessing the robustness of neural networks.

**ISO/IEC TR 24028:2020: Overview of trustworthiness in artificial intelligence**
It describes approaches to building trust in AI systems through transparency, explainability, controllability, among other criteria, technical pitfalls and typical threats and risks associated with AI systems, as well as possible mitigation techniques and methods, and approaches to assess and achieve the availability, resilience, reliability, accuracy, safety, security, and privacy of AI systems.

**ETSI SAI 005: Securing Artificial Intelligence (SAI); Mitigation Strategy Report**
In line with the ETSI plan[70], this standard summarizes and analyses existing and potential mitigation against threats for AI-based systems.

**IEEE P7002 - IEEE Draft Standard for Data Privacy Process**
P7000 series, looks at data ethics in systems development. It addresses how organizations should:

- manage privacy risk;
- identify their privacy requirements;
- develop and manage systems;
- meet their privacy requirement.

**ISO/IEC AWI 5259: Data quality for analytics and machine learning (ML)**
This is an Approved Work Item to define requirements and guidance for establishing, implementing, maintaining, and continually improving the quality for data used in the areas of analytics and ML.

**ISO/IEC 42001:2023 - Information technology - Artificial intelligence - Management systems**
**Overview**
ISO/IEC 42001:2023 provides a framework for establishing, implementing, and improving an Artificial Intelligence Management System (AIMS) to ensure ethical governance, risk management and accountability in organizations using AI. It is applicable to organizations of all sizes and sectors involved in the development or deployment of AI systems.

**Development Process**
The standard was developed through collaboration among experts in technology, ethics, law, and business to address the complex challenges posed by AI and to ensure alignment with ethical and regulatory standards.

[70] Artificial Intelligence and future directions for ETSI 1st edition – June 2020 ISBN No. 979-10-92620-30-1
https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp34_Artificial_Intellignce_and_future_directions_for_ETSI.pdf


© Copyright SMPTE 2023, 2025 – All Rights Reserved

**Key Components**

- AI Management System (AIMS): Integrates with organizational processes for continuous improvement and alignment with other ISO standards.
- Risk and Impact Assessment: Assesses the risks and potential impacts of AI systems on individuals, organizations, and society.
- Privacy and Security: Ensures compliance with privacy laws and protects AI systems from threats.

**Structure and implementation**
The standard follows the structure of other ISO management systems and integrates with standards such as ISO 9001 (quality) and ISO 27001 (information security). Implementation steps include conducting a gap analysis, developing an AIMS, performing risk assessments, and preparing for certification.

**Key objectives**

- Establish clear policies and procedures for AI systems.
- Ensure transparency and accountability in AI decision making.
- Mitigate bias and ensure privacy and data security.
- Align AI practices with ethical principles and regulatory requirements.

**Controls**
The standard includes controls organized into nine objectives, covering areas such as AI governance, lifecycle management, data security, and third-party relationships.

### 13.7 Data-related Standards

In recent years, AI has moved from a model-centric approach to a data-centric approach. The quality of AI applications is heavily influenced by the quality and relevance of the training data. This means that the data must be properly annotated using a semantic approach.

The feasibility of using large datasets across industries will open the market to more innovation in AI. Standards for the safe, efficient, and reliable exchange of information within datasets will be one of the most important developments for AI in the coming years. ETSI has work streams dedicated to the data supply chain and availability of training data, while ISO has several projects on AI and big data.

## 14 Opportunities for new AI/ML standards

### 14.1 Overview

The motivation behind standards in general is a desire for improving interoperability between implementations or systems. Much of the AI/ML ecosystem is descended from open source frameworks. The nature of open source systems is that maintainers can regulate contributions from the community to evolve tools and systems, which allows for rapid (or slow) evolution based on the preferences and needs of the maintainer(s). End users can choose to stick to a particular version of a tool or framework or migrate as new features are released. The open source model obviates the need for certain types of standards, as it allows tools and frameworks to evolve in response to user needs, bugs, and contributions rather than due process. However, consensus among stakeholders is not a mandatory requirement among open source developments.

© SMPTE 2023, 2025 – All Rights Reserved

Moreover, many large organizations have deployed tools and services built on internally developed frameworks and processes. These tools and services are often offered as turnkey AI/ML solutions, so the main interoperability concerns are around the inputs and outputs of the tools, which can often be specified in a product datasheet. Such organizations might view standards at best as unnecessary and potentially even as harmful to business interests and competitive positioning.

Despite this landscape, standards and recommended practices can still serve an important role in the AI/ML ecosystem. The following subsections describe some areas where such documents might be valuable.

## 14.2 Ontologies

Various groups have defined a number of ontologies for defining relationships within media. Examples include Ontology for Media Creation[71] (MovieLabs) and EBUCorePlus[72] (EBU). Studios also have their own internal ontologies. This fragmentation has necessitated the development of tools to "translate" from one ontology to another.

A potential standard that might be useful would be one that quantifies how to measure the "quality" of ontology, particularly as it relates to specific use cases, so that content creators have an idea of how to evaluate and properly deploy ontologies. One approach could be to provide guidelines and templates for designing ontologies for specific purposes. Another method would be to share application-specific data and evaluate the performance of applications implementing ontologies.

## 14.3 Model Metadata

A standard for model metadata and/or dataset metadata might be very helpful for ensuring that model and dataset characteristics can be understood across implementations. Vendor-specific methods exist for tracking certain types of metadata (e.g., training parameters, evaluation metrics, dataset versions, etc.); however, these can vary from implementation to implementation.

For example, attributes such as model revision, time stamp, framework version, etc., could be useful to standardize. Dataset metadata such as revision, demographic content, license/usage information, etc., might also be helpful. For auto-tagging applications, it might be helpful to standardize methods to describe how metadata was generated (auto/manual and versioning model) and perhaps how it is intended to be used. It would be helpful to document how to measure "metadata quality."

The datasets used for training must be labeled and versioned for machine learning applications. Indeed, these datasets greatly influence the quality of the trained models and the application. An AI model trained with biased data will be biased at the end. It can occur, for example, in face recognition systems. Suppose the deep learning model in the pipeline has been trained mainly with Caucasian male faces. In that case, the application's performance will be inconsistent based on the gender and ethnicity of the personalities to be recognized. The metadata attached to content for training will facilitate the process to evaluate the bias a priori (i.e., before training the models). In many cases, a training set is extracted from a large dataset. Having structured metadata attached to the content used for training will significantly facilitate the reusability, merging, and enrichment of resulting datasets.

---

[71] Ontology for Media Creation documentation and resources are available at: https://mc.movielabs.com/docs/ontology/

[72] EBUCorePlus documentation and resources are available at https://github.com/ebu/ebucoreplus

 © Copyright SMPTE 2023, 2025 – All Rights Reserved

The generative AI era introduces complexity beyond traditional model metadata, as modern AI systems perform diverse tasks by modifying inference parameters such as prompts or context, making model identifiers alone insufficient for ensuring reproducibility and interoperability in media workflows. Recognizing this challenge, a SMPTE TC-30MR drafting group (30MR DG AI Model Metadata) is developing a comprehensive metadata schema and identifier system for AI systems configured to perform specific media-related tasks.

This approach treats each deployment as a unique, task-specific AI configuration. Each configuration encompasses the underlying model along with its operational parameters, prompts, contextual settings, and input specifications. Each configuration is described through a structured metadata record that includes identification and comprehensive task descriptions, configuration parameters, and essential administrative information, including training data summaries, licensing terms, and ethical constraints.

This framework ensures traceability, interoperability, discoverability, and reproducibility across complex media workflows through specific applications, for example: making embeddings understandable and interoperable by clearly describing the embedding generation process and parameters; ensuring LLM-based metadata generation remains reproducible across different production environments; and guaranteeing voice cloning for dubbing maintains consistency throughout film and series production cycles. These capabilities provide the granular identification necessary for modern, AI-driven media production environments where consistent, repeatable results are essential for professional content creation.

## 14.4 Benchmarking

The ML Commons organization has defined some well-accepted benchmarks for generic ML training and inference on tasks such as machine vision and recommendation engines; however, media-specific benchmarks for common tasks (such as those in Table 3) using standardized testing datasets would be helpful for creators to quantify performance of systems. The EBU has defined a benchmark for speech-to-text and is developing a benchmark for facial recognition. While systems can certainly be defined in the absence of a standard benchmark, benchmarks can allow for a fair comparison of systems using datasets and metrics agreed by end users as being important.

**Table 3 — Opportunities for AI/ML benchmarking standards**

| **Common uses of AI/ML in media that might benefit from benchmarks** |
|---|
| ***Video:*** |
| Fidelity enhancement or restoration |
| Content identification (e.g., detecting logos, products, piracy, etc.) |
| ***Audio:*** |
| Automatic closed captioning |
| Machine translation/localization |
| Speech-to-text |
| ***Metadata:*** |
| Automatic tagging of missing metadata |
| Metadata quality control |

© SMPTE 2023, 2025 – All Rights Reserved

**Benchmarking of LLMs**
Work suggestions for Standards:

- Define elementary tasks based on media-specific use cases to evaluate the models with simple metrics.
- Write a guide on how to evaluate models fairly on complex and well-defined tasks (for journalists, scriptwriters, editors, engineers...), this guide implies subjective human evaluation.
- Compare open source and commercial solutions for these specific tasks and share the results as done by research centers.

## 14.5 Recommender Systems

Standardization efforts can be directed at the following points:

- **Promote transparency and accountability:** Establish clear criteria for the type of data that can be used by AI recommendation systems. This should include standards for respecting user privacy and complying with regulations, such as GDPR. There should be transparency about the data being collected and the purpose of its use.
- **Establish evaluation metrics:** Standardize ways to measure the performance of recommendation systems. This would allow different systems to be compared on a level playing field and encourage improvements in these key areas.
- **Establish guidelines for addressing bias and the bubble effect:** Guidelines should be established to mitigate bias in recommendations and to address the bubble effect.
- **Establish a certification process:** A formal certification process could help ensure that recommendation systems meet these standards before they are released.

## 14.6 Data Usage Recommended Practices

Many organizations would find it useful to refer to guidelines on best practices for using data, particularly data that might be scraped from the internet or without a clear license. There is not a general consensus as to whether model parameters (biases and weights) that are trained using copyrighted material are considered to be a derived work for the purposes of copyright law. Developing a Recommended Practice on this subject would require considerable input from intellectual property experts and lawyers. The Text and Data Mining (TDM) exception is not straightforward in the Digital Service Market (DSM), cf. article 4:

https://en.wikipedia.org/wiki/Directive_on_Copyright_in_the_Digital_Single_Market

This exception applies to content subject to copyrights for research purposes only. But it seems that this exception covers only research entities recognized as such for using content subject to copyright. But for other entities, the owners of rights can opt out of the TDM exemption. The TDM exception is broader and not restricted to research applications in the US. The EBU is considering the implications of using content from archives that is subject to copyright to train models.


© Copyright SMPTE 2023, 2025 – All Rights Reserved

## 14.7 Cloud Computing

There are a diverse set of APIs for different cloud providers, and it is not easy to utilize a hybrid-provider cloud infrastructure. Many companies maintain their own data and compute resources, and there does not seem to be a strong desire to standardize APIs across providers. Some cloud providers' policies regarding data usage allow data to be used for purposes beyond users' immediate business needs (e.g., to train or improve models whose use could be sold to competitors or others).

For production workflows, the use of services from multiple service providers and multiple cloud providers is a challenge. The SMPTE 34CS Technology Committee, in collaboration with the EBU MCMA Working Group, has developed a cloud-agnostic framework that abstracts the specificity of services to facilitate the reuse and integration of these services in complex workflows.

## 14.8 AI Ethics

Many media-specific areas of interest exist with regard to AI ethics: privacy; AI-enabled interactions, such as with virtual characters; censorship; diversity; bias; accountability; etc. Documenting ethical best practices for AI systems in media use cases would be welcomed by many organizations. Considerations around AI ethics are discussed further in Section 12.

## 14.9 MCP and A2A

MCP (see Section 9) is not currently a de jure standard. Formal MCP standardization could enable media technology providers to focus on innovation within their specialized domains while ensuring all components in the media production chain work harmoniously toward delivering exceptional content experiences. From a standardization perspective, the Model Context Protocol represents the interoperability layer our media technology ecosystem has long needed—creating a unified framework where production, post-production, and distribution systems communicate seamlessly rather than through today's complex web of proprietary integrations. For instance, an AI-powered script analysis tool could use MCP to seamlessly retrieve shot metadata from a production management system, automatically suggest VFX requirements to the visual effects department, and simultaneously update budgeting software—all through standardized tool calls rather than custom integrations for each system pair.

Similarly, the A2A (see Section 9) framework is also not a de jure standard. A2A standardization could allow media technology providers to abstract away the complexities of inter-model communication, enabling better interactions between agents.

Just as standards such as SMPTE ST 2110 revolutionized IP-based media transport, the combination of MCP for tool interoperability and A2A for agent collaboration could form the backbone of modular, future-proof workflows where specialized applications can be added or upgraded without disrupting the entire production technology stack. This approach would ultimately free creative professionals from technical integration concerns, enabling them to select best-of-breed tools that automatically connect through a common protocol, thereby accelerating both technological advancement and creative possibilities.

If MCP and A2A are not standardized, implementers might face unexpected challenges around interoperability, security, and/or licensing.

© SMPTE 2023, 2025 – All Rights Reserved

# 15 Datasets and the Need for Data

## 15.1 The Importance of Data

An AI model is only as good as the data that is used to train it. The public availability of large, annotated datasets has enabled rapid advances in performance of AI systems. For example, with such datasets, researchers can engage in competitions to develop optimized AI models for specific domains (e.g., Kaggle). Public datasets are also useful for early prototyping and development of AI models.

Development and curation of large datasets can involve a large amount of effort and cost. Many types of data are publicly available on the internet, but media is generally copyrighted, and it is often unclear whether the owners would permit the use of such media for AI model development. In addition, supervised learning models require annotations that can be labor-intensive to create and QC.

## 15.2 Public Datasets

Table 4 lists just a few of the noteworthy AI datasets related to media.

**Table 4 — Noteworthy datasets related to media**

| Name | Domain | Size | License | Website |
|---|---|---|---|---|
| FineVision | Vision Language Models | 5TB of images and text from 200 datasets | Per original dataset and CC-BY-4.0 for Hugging Face contribution | https://huggingface.co/spaces/HuggingFaceM4/FineVision |
| ImageNet | Object detection/ recognition | 14,197,122 images | Unknown* | https://www.image-net.org |
| Open Images V7 | Object detection/ recognition | 9,178,275 images | CC BY 2.0<br>CC BY 4.0 | https://storage.googleapis.com/openimages/web/index.html |
| MS COCO | Object detection/ recognition | 330,000 images | CC BY 4.0 | https://cocodataset.org |
| YouTube-8M | Entity recognition | 6,100,000 videos | CC BY 4.0 | https://research.google.com/youtube8m |
| YFCC-100M | Metadata | 99,171,688 images/ 787,479 videos | CC (various) | https://multimediacommons.wordpress.com/yfcc100m-core-dataset/ |
| IMDB | Movie reviews | 50,000 reviews | Noncommercial use | https://www.imdb.com/interfaces |
| Common Voice | Speech recognition | 24,211 hours | CC0 | https://commonvoice.mozilla.org/datasets |
| People's Speech (in development) | Speech recognition | 100,000 hours | Open | https://mlcommons.org/peoples-speech |
| Million Song dataset | Music | 1,000,000 songs | Various | http://millionsongdataset.com |
| SLLFW<br>CALFW<br>CPLFW | Face recognition | 3,000 pairs each | Unknown | http://www.whdeng.cn/SLLFW<br>http://www.whdeng.cn/CALFW<br>http://www.whdeng.cn/CPLFW |
| CFPW (celebrities) | Face recognition | 7,000 images | Unknown | http://www.cfpw.io/ |


© Copyright SMPTE 2023, 2025 – All Rights Reserved

| Name | Domain | Size | License | Website |
|---|---|---|---|---|
| VGG-Face2 | Face recognition | 3,310,000 images | Unknown | https://github.com/ox-vgg/vgg_face2 |
| TVSum | Video summarization | 50 videos | CC BY 3.0 | https://github.com/yalesong/tvsum |
| SumMe | Video summarization | 25 videos | Unknown | https://zenodo.org/records/4884870 |

* Use requires submitting a form certifying non-commercial

## 15.3 Need for Future Datasets

Several media organizations identified a need for standardized datasets with permissive license policies to help them develop and evaluate AI models. This could involve sponsoring the creation of suitable academic datasets or pooling contributions from various organizations under a permissive license.

Some companies are interested in benchmarking models using a standardized set of copyrighted content, but it is difficult to convince rightsholders to license content for this use.

Standardized datasets for benchmarking are needed in the following areas:

- Captioning
- Transcription
- Speaker and celebrity identification

## 15.4 Alternatives to Public Datasets

Because of the difficulties involved with licensing content for public use for ML applications, a possible alternative approach would be the creation of private datasets, where rightsholders could contribute content towards a specific usage. For example, the EBU has released a face recognition dataset containing more than 80 hours of annotated TV programs, which is available upon request. An option to facilitate the sharing of data for testing and training models is to use the federated learning approach. The principle of federated learning is to move the models to the data instead of moving the data to the models. This approach is conducive to standardization because datasets must be strictly structured to be usable by machine learning models. One key advantage is that the data can be kept private while allowing models to be trained across organizations. However, this approach requires close collaboration between organizations and limits opportunities for differentiated models since the trained models would be shared.

© SMPTE 2023, 2025 – All Rights Reserved

# 16 Conclusion

The following is a summary of the Task Force's findings:

- AI is a foundational and transformative technology that is still in its early stages. There are expected to be many step-change advances that take place in the coming years. AI is already massively disrupting certain industries (such as software engineering). Over the longer term, the cumulative effect will undoubtedly disrupt the entire media industry, from creation to consumption.

- There are many groups that are exploring how standards might be able to improve interoperability and reduce friction to developing and deploying AI. Private entities have proposed foundational frameworks such as MCP and A2A that merit formal standardization.

- AI advances are often driven by the widespread availability of large datasets to help researchers and developers hone their models. Media-specific datasets, particularly those for benchmarking specific tasks, would benefit the industry, and companies should explore ways of enabling the creation of such datasets.

Considerations around ethics are paramount when developing and deploying AI systems. Because AI systems are constantly performing new and unexpected tasks, many people have discomfort around the future role of such systems and societal implications. AI systems must be built around principles of trust, fairness, and inclusion. As stated consistently in these pages, there is no real playbook for AI ethics. The technology is early, and its ethical considerations are lagging behind the pace of innovation. Even regulators have, by and large, limited their interest to data privacy—for now. But because the technology is so complex and increasingly important, and its presence is so ubiquitous, it must be approached and roadmapped through multifaceted systems thinking. There are simply too many components in any serious artificial intelligence effort to avoid considering its requirements and ramifications—especially ethics—as anything but a system. The most successful organizations in building and scaling AI internally are those that think about it the most thoroughly and systematically. And nothing forces an organization to think deeply—and in systems—more than ethics. Far from window-dressing or virtue signaling, putting ethics front and center will bring about the modes of operation, intellectual rigor, and organizational culture necessary to excel in building AI systems.

# 17 Acknowledgments

SMPTE would like to thank the EBU and the ETC for their major contributions to this work. This report would not have been possible without the efforts of the members of the task force who participated in drafting this report:

| | | | |
|---|---|---|---|
| Alex Rouxel | Alex Zou | Andy Maltz | Andy Rosen |
| Benjamin Ing | Bill Redmann | Brian Schunck | Brian Vessa |
| Bruce Devlin | Chaitanya Chinchlikar | Chris Lennon | Dean Bullock |
| Ellen Ryan | Fred Walls | Gheorghe Berbecel | Jian-Rong Chen |
| Jim Helman | JoAnne Kim | Jonathan Thorpe | Karen Kilroy |
| Lars Borg | Leigh Whitcomb | Nicole Quirk | Olga Howard |
| Paul Gardiner | Paul Treleaven | Pete Sellar | Radoslav Markov |
| Raymond Yeung | Scott Smyres | Spencer Stephens | Stephen Scott |
| Steve Llamb | Sujay Kumar | Thomas Bause-Mason | Tryc Wojtek |
| Veronica Pineda | Will Kreth | Yves Bergquist | Zaidan Alaoui |


 © Copyright SMPTE 2023, 2025 – All Rights Reserved