Associated Press

Artificial Intelligence

To accurately explain AI technologies to their audiences, writers must understand both the vocabulary associated with AI technologies (see below Key AI terms) and the concepts used in their creation. The following guidance also can serve writers seeking to use AI models in their coverage.

GENERAL GUIDANCE for AI COVERAGE

Stories about specific AI technologies should explain how they were built, as well as the mechanics and basic makeup of the AI system. Writers should seek to evaluate the data used to train the model, the outputs the model generates, and the model itself, as critically as they would any other source of information. That's important whether the tool is used to guide self-driving cars, facilitate speech recognition, generate written text for a chatbot, or any other use.

Stories about AI technologies should describe what, specifically, such tools are built to predict or to optimize, whenever that is possible. Those details will help reporters learn which applications of the model may be most appropriate, as well as how the original intent may have influenced its design.

Broadly, many such systems are created to identify patterns in large volumes of data, then to make predictions based on those patterns. For example, if we think of a simple algorithm as being like a recipe, composed of a detailed set of instructions that describe how to solve a problem or create a certain dish, we need to first understand how the problem or end goal originally was defined.

Many chatbots, meanwhile, are built using a complex, multilayered design that is meant to predict which words and phrases are most likely to appear together — but is not meant to generate human-level comprehension.

Knowing as much as possible about the rules that guide an AI system, in turn, can reveal more about how well the system works, or what engineers often call the "accuracy" of an AI system.

Stories about AI technologies should explore any outside checks on the design and deployment process. That includes the phases of data acquisition, creation, sharing and sales, as well as testing, auditing and tracking.

Stories about AI technologies should be explicit that these systems are developed, built and deployed by humans, even if the most advanced models can later train themselves without our intervention. As such, they reflect human choices and risk reflecting human biases, which could be present in the data that powers the model as well as the rules that guide it. They are also dependent on human labor, from engineers working in corporate labs to independent wage laborers hired to label the data powering certain models.

Stories about AI benefit greatly from interdisciplinary collaboration at the intersection of technology, ethics, social science and governance to understand these tools holistically. Consulting with academics, policy experts and outside reviewers allows reporters to enact an independent check on such tools and tease out their promise and their limitations. Such conversations also allow reporters to bulletproof their own conclusions if attempting to reverse-engineer or test these tools themselves.

Stories about AI should take into account the larger societal context in which these systems are deployed. That includes seeking out thoughtful conversations with communities who may be negatively affected by such tools, which may vary widely depending on the type of tool and context.

Here are some suggested reporting approaches for writers covering these technologies and their effects:

- Highlight the people who develop, build, implement and test these systems. Tracing the human origins of AI systems will help unlock new reporting lines.
- Focus reporting on communities that may be negatively affected, particularly by surveillance and predictive technologies, and how such tools can magnify historical biases that may be present in the training data or in the model's construction or application.
- Include context about the limitations and risks of such tools in the specific social contexts in which they are applied.
- Explore transparency issues surrounding the deployment of AI systems. Do chatbot users have visibility into the primary source material that makes up the text responses the tools produce? Or, for example, if an algorithm is used to make decisions for a government agency, can a person see the risk score they were assigned by that algorithm?

Here are some common pitfalls to avoid:

- Many AI developers describe their tools as breakthrough or revolutionary technologies, but few such systems truly are. Beware of far-fetched claims that bear more resemblance to thinly disguised marketing ploys.
- Narratives about AI systems wiping out humanity have been around for decades. While enticing to explore, these risk distracting attention from real-world problems that such systems can cause in the present day. Rather than focusing entirely on far-off futures, focus on current-day concerns about such tools' reliability, replicability and validity, as well as their potential for bias.
- Do not ascribe human emotions or capabilities to AI models or imply that they exert independent, human-like agency in the world. It does not further the public's understanding of their role. Instead, it obscures humans' work in developing and supervising such AI systems.
- Do not illustrate every piece of journalism about AI with an image of a robot, or a humanoid machine. Such imagery can give the false impression that AI systems are embodied, rather than pattern-identifying systems made from discernible data inputs.

Here are some more technical questions to ask when reporting on the inner workings of AI tools:

- What variables were selected to build the model? What weights were those variables assigned? When possible, request the data fields, the types of data they contain, details about any codes in the data and how missing data is indicated.
- What type of AI model was developed? What is the model optimizing for, or predicting?
- What use case was the model designed for? Has it been applied in other contexts?
- What training data was used to guide the AI tool? Is that data representative? Is it up to date?
- What other systems or technologies does the tool integrate? For example, self-driving cars may rely on radar analysis and computer vision systems both built on training data.
- How is the model's performance being evaluated? According to those evaluation criteria, how well did the model perform? Was that better than the human error rate?
- How were humans involved in the process when the tool was built? When it was first deployed? As it is now used in the real world?
- Who is impacted by the model? If applicable, what people or communities are negatively affected by the model and specifically, how?

KEY ARTIFICIAL INTELLIGENCE TERMS

Writers are uniquely positioned to help our audiences grasp when artificial intelligence models power scientific breakthroughs, and when they help spread hate or fuel structural discrimination. Even if newsrooms lack the tools to reverse-engineer an AI tool to explain its specific role in guiding decisions in hiring, banking or medicine, understanding key terms will help guide crucial journalistic questions. Writers should become familiar with these terms so they can accurately and succinctly explain the societal changes that AI and generative AI models can mean — including their potential, their inherent risks and their varying effects on different groups.

Some key points, as noted in the first definition below: **Avoid language that attributes human characteristics to these systems, since they do not have thoughts or feelings but can respond in ways that give the impression that they do. Do not use gendered pronouns in**

referring to AI tools. And keep in mind that such systems are built by people who have their own human biases and aims.

artificial intelligence

Computer systems, software or processes that can emulate aspects of human work and cognition. Such systems are not conscious but are trained on vast datasets to accomplish tasks such as visual perception, analyzing and using speech, and learning — although in many cases, only to a limited extent. The term itself has been the subject of debate over the definition of intelligence as the technology has evolved in scope and influence.

AI is acceptable in headlines and on second reference in text.

Avoid language that attributes human characteristics to these systems, since they do not have thoughts or feelings but can respond in ways that give the impression that they do. Do not use gendered pronouns in referring to AI tools. And keep in mind that such systems are built by people who have their own human biases and aims.

The terms artificial intelligence and artificial general intelligence are not synonymous.

artificial general intelligence

An emerging branch of artificial intelligence that aims to build AI systems that can perform just as well as — or even better than — humans in a wide variety of tasks, including reasoning, planning and the ability to learn from experience.

Some developers say the technology could result in broadly intelligent, context-aware machines that could adapt to be used effectively in a variety of settings. However, other researchers say that it would take a very long time for such systems to achieve “human-level” intelligence, which relies on inherent human traits such as sensory perception, creativity, understanding emotion and critical reasoning.

Right now, AI systems can emulate only aspects of human work and cognition, but are not sentient per se.

The definition of artificial general intelligence, and concepts about how it differs from human intelligence, have changed over the years. When evaluating claims of artificial general intelligence, consider the source and their motivations and beware of so-called breakthroughs, because few things truly are.

Don’t use AGI on second reference unless necessary in a direct quotation, in which case explain the term.

The terms artificial intelligence and artificial general intelligence are not synonymous.

Algorithm

Detailed computational instructions that describe how to solve a problem or perform a specific task. A simple, real-life example is a recipe, which describes both a set of inputs — i.e., ingredients — and an output consisting of the dish itself. Machine learning algorithms are tools that are “trained” with large datasets to improve the predictions they generate. For example, some algorithms try to predict which posts a person wants to see on a social media platform, while other algorithms dispense targeted recommendations to each user who visits a shopping website. Some of the most highly complex machine learning algorithms are not always fully understood by their creators.

Algorithmic bias, which are decisions guided by AI systems that result in discriminatory or disproportionate effects on certain groups of people, has emerged as a major issue raised by critics and government agencies alike.

algorithmic bias/AI bias

Decisions guided by AI tools that replicate and amplify human biases, leading to discriminatory outcomes that can systematically impact specific groups of people.

Algorithmic bias is often used to describe the negative impacts of tools that draw from large datasets that are skewed by historical or selection bias.

For instance, an AI sentencing tool trained on historical data showing that Black offenders typically received longer prison terms than white offenders for the same crimes could yield predictions that incorporate that bias — in effect, reproducing past injustices.

In another instance, U.S. civil rights agencies have found that some computer-based tools for hiring workers or monitoring their performance may disadvantage people with disabilities.

Such bias can occur along lines of age, race, color, ethnicity, sex, gender identity, sexual orientation, religion, disability, veteran status, class and many other variables.

These technologies also can perpetuate AI bias, a different class of errors that can result from the creation or use of AI tools. For example, AI bias can arise as a result of decisions made in the design of the AI tool or the historical models it draws from, or in the societal context in which humans use the AI system.
Explain either term if writing for general audiences.

ChatGPT

An artificial intelligence text chatbot made by the company OpenAI that was released in late 2022. It relies on technology known as a large language model, which is trained to mimic human writing by processing a large database of digitized books and online writings and analyzing how words are sequenced together. People can ask ChatGPT — and similar chatbots made by the company's many other competitors — to answer a question or generate new passages of text, including poems, letters and essays. It responds by making predictions about what words would answer the prompt it was given.
Tools such as ChatGPT show a strong command of human language, grammar and writing styles but are sometimes factually incorrect. Avoid language that attributes human characteristics to these tools, since they do not have thoughts or feelings but can sometimes respond in ways that give the impression that they do. Like other AI models, ChatGPT can be prone to algorithmic bias that may skew its responses and analyses. Outside researchers' inability to probe its training data set also complicates efforts to understand how it settles on its responses, what information it relies on and how it reaches conclusions.
ChatGPT's popularity after its release helped spark public fascination and commercial interest in similar technologies, and numerous competitors also have introduced their own chatbots built with large language models. Some companies, including Google, Microsoft and startups, have released their own publicly accessible chatbots, while others use the technology internally or sell it directly to businesses.
Some but not all commercially available chatbots are powered with GPT, which is an abbreviation for generative pretrained transformer. Use chatbot as the generic term. Don't use GPT or ChatGPT to refer to all chatbots.

effective altruism

A social and philanthropic movement popular in Silicon Valley that has influenced the trajectory of artificial intelligence. Followers say they try to maximize the good they can do in the world and have funded philanthropic efforts to protect against what they see as the worst potential future outcomes of artificial intelligence, among other issues. Sam Bankman-Fried, the now-convicted founder of the cryptocurrency exchange FTX, was among the most visible proponents of effective altruism. Explain the term when used.

face recognition

A technology for automatically detecting human faces in an image and identifying individual people. It is a form of biometric technology that relies on comparing aspects of a face against a database of images to find a match. Techniques for comparing facial features to recognize individual faces have existed since the 1960s, but the technology has improved through advancements in computer vision, machine learning and data processing.
Face recognition raises privacy and accuracy concerns because governments and others can scan images from video cameras or the internet and track individual people without their knowledge, and some systems have been shown to work unevenly across demographic groups. Some lawmakers have sought to curtail the technology as it becomes more widely used by law enforcement, businesses and consumers.
Similar technologies include gait recognition, for detecting people in video images based on their body shape and how they move; and object recognition, for detecting objects in an image, such as a traffic cone in the path of a self-driving car.
Face recognition technology is sometimes called facial recognition technology or face scanning.

generative AI

A term for AI systems capable of creating text, images, video, audio, code and other media in response to queries. Humans can interact with generative AI models in a seemingly natural way but the models aren't reliably capable of distinguishing between facts and falsehoods or fantasy. Generative AI systems often are powered by large language models. They sometimes generate inaccurate or fabricated responses to queries, an effect AI scientists call hallucination. If using the term hallucination, describe it as an issue associated with the technology that produces falsehoods or inaccurate or illogical information. Some in the field prefer the term confabulation or simpler terms to describe the inaccuracies that don't draw comparisons with human mental illness.

large language models

AI systems that use advanced statistics to uncover patterns in vast troves of written texts that they can apply to generate responses. Such systems are increasingly capable of applying the syntax and semantics of human speech and can also be used to generate a variety of media. The models work based on the probability that certain words and phrases appear together, and their level of sophistication and accuracy can vary across human languages. GPT, an AI system created by the Microsoft-backed company OpenAI, is a large language model. Do not abbreviate as LLM outside technical contexts.

machine learning (n., adj.)

An AI process in which computer systems identify patterns in datasets to make or refine the decisions and predictions that they generate without being explicitly programmed to do so. Examples of machine learning applications include face recognition, language translation and self-driving cars.

Explain the term if writing for general audiences. Machine learning is not a synonym for AI.

training data

A dataset used to teach an algorithm or a machine learning model how to make predictions. Because the models learn to find patterns from the training data, it is important to consider the specific information it may contain. The types of training data used in different AI tools can vary widely, from large quantities of written texts to vast digital libraries of images of human faces to historical arrest records from specific geographic areas.