



## Education in the Time of AI<sup>1</sup>

---

The term “artificial intelligence” was coined in 1955, in a proposal written by John McCarthy, Marvin Minsky, Nathaniel Rochester and Claude E. Shannon for what became the Dartmouth Summer Research Project on Artificial Intelligence—widely considered the founding event for the field. The project, they explained, would “proceed on the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves.”

Designing, building, and improving computer systems capable of performing tasks normally associated with human or animal intelligence remains a core goal of AI research. AI applications have already become part of the machinery of modernity and are likely to become more sophisticated and essential to our institutions, industries, and individual selves.

In education, AI offers potentially profound opportunities as well as serious concerns. We cannot know yet exactly what effects it will have on students, staff, and faculty, but it will certainly affect life at Rhodes within and beyond the classroom.

### Is It Friend or Foe?

It can be either or both, depending on user and use.

There are broad ethical implications involving attribution, permissions, copyright, and privacy to consider when using it for any reason: OpenAI, for example, scrapes copyrighted material from which it profits and for which it does not pay; it collects information about its users (IP addresses, data in interactions, browsing activities); and it may share information with unspecified third parties. This, of course, does not distinguish it from much of the internet.

---

<sup>1</sup> Prepared by Dr. Seth Rudy and the Open AI/Chat GPT Task Force

Ethical issues are also being worked out by share- and stakeholders with potentially competing or conflicting goals: for instance, Microsoft’s layoffs earlier this year included “its entire ethics and society team within the artificial intelligence organization.”<sup>2</sup> The need for such teams, and the extent to which they are or are not adequately supported, may raise additional questions about the ethics of AI as a technology in and of itself.

AI is also a double-edged sword in terms of environmental impact. It may help find or implement solutions to the climate crisis, but it is itself a consumer of resources with a large carbon footprint. Researchers have estimated that even before GPT-4, ChatGPT used up “500 ml of water – a standard-sized water bottle – every 20 questions and corresponding answers.” Future iterations are likely to be even thirstier.

The picture on campus is no less complicated. The potential for plagiarism in traditional essays is certainly high; AI can also write code and assist with data analytics, but (as with essay composition) only to a relatively simple degree of sophistication or reliability. This too is likely to change. A recent study from Stanford found that ChatGPT has become less accurate over time, but not everyone is convinced by the findings, and it has been reported that it can already pass freshman year at Harvard.

If some are alarmed by its potential to hinder learning and facilitate cheating, though, then others are excited by its ability to assist with generating ideas, providing useful information, and improving efficiency. It may also be helpful to students with academic accommodations for ADHD, dyslexia, etc. Students might reasonably be concerned that not understanding or having facility with AI (E.g., “prompt engineering”) may adversely affect their employment opportunities: as one Rhodes student reported, “if there’s a way to use it for something that isn’t writing a paper, but could help in a job setting then perhaps we should consider ways in which to use it, otherwise we could fall behind.”

## What AI Systems and Tools Are Out There?

There are many web-accessible sites and systems offering users AI-powered “assistance” with cognitive tasks, and their numbers continue to increase. What follows are descriptions of a very few noteworthy or representative examples; by the time you read this, some may have been acquired by other companies, integrated into other platforms, or ceased to exist. [Futurepedia](#) maintains a comprehensive and continually updated directory of AI tools.

**Aceify.ai:** Aceify markets itself as a new “study companion.” It works as a Chrome extension that can gather information about online quizzes and assessments and select the correct answers (specifically to multiple-choice, fact-based questions) at the click of a

---

<sup>2</sup> The company still has an active Office of Responsible AI.

button. The extension is compatible with several Learning Management Systems, including Canvas.

**ChatGPT:** in its own AI-generated words, ChatGPT “is a conversational artificial intelligence language model developed by OpenAI. GPT stands for ‘Generative Pre-trained Transformer,’ and it is part of the GPT series, which includes several iterations...It is a Large Language Model (LLM) based on the Transformer architecture, which enables it to process and generate human-like text by predicting the next word in a sequence given a context of preceding words.

ChatGPT is specifically designed to have interactive conversations with users. It can understand and generate human-like responses, making it capable of having relatively natural and coherent conversations on a wide range of topics. The model is pre-trained on a vast corpus of text from the internet, which allows it to gain a broad understanding of language and knowledge. In other words, ChatGPT is an AI that can generate high-quality compositional and informational content in response to human prompts, but may also produce unreliable rubbish or anything in between.

This is true, *for now*, of most if not all other text-generative AI tools based on LLMs, other examples of which include Claude, Bard, Bloom, Cohere, Falcon, Jasper, LaMDA, LLaMa, Orca, Sudowrite, and more.

**Cramly:** A writing “assistance” service, this site promises students “higher ROI on their education, by maximizing the value of every student's study time.” Highlights include an “outline generator” and “plagiarism checker,” “just in case you think like the greats.” Like Grammarly, but cheaper.

**Dall·E:** Another OpenAI system, Dall-E is a text-to-image generator operable from within ChatGPT. As explained by OpenAI, “when prompted with an idea, ChatGPT will automatically generate tailored, detailed prompts for DALL·E 3 that bring your idea to life. If you like a particular image, but it’s not quite right, you can ask ChatGPT to make tweaks with just a few words.” Results are often striking, precise, and can appear photorealistic.

Intellectual property and other potential legal concerns have led OpenAI to introduce mitigation factors regarding harmful content, public figures, misinformation, and representation.

**Grammarly:** Launched in 2009 and now widely used by students, Grammarly is a suite of AI-powered “writing assistance” tools that promises help with brainstorming, composition, citation, and revision. The free version comes with 100 prompts for AI text generation; the premium version includes tone adjustment, sentence rewriting, and “accidental plagiarism” detection features, and comes with 1,000 AI prompts.

**Midjourney:** Like DALL-E 3, Midjourney is an AI-powered text-to-image visual art generator that uses neural networks trained on large datasets of images. The service is currently only accessible through a Discord bot on their official Discord server.

**Quillbot:** Quillbot offers a selection of writing “assistance” tools using Natural Language Processing (NLP) techniques. Users can use the software to paraphrase or summarize texts, check grammar and spelling, and generate content. There are stylistic options as well: users writing in English can select “Formal,” “Academic,” “Simple,” and “Creative” output versions, or click buttons to “Expand” or “Shorten” passages. Other languages offer only the “Standard” mode of paraphrase.

**Stability.AI:** This open-source generative AI company is currently best known for its Stable Diffusion deep-learning, text-to-image model. The company has also launched diffusion models for AI-generated audio and video.

## Can We Detect Them?

Not computationally—or least, not reliably. Detection tools do exist: Turnitin.com, ZeroGPT, and others suggest their programs can distinguish between human and AI-written content, but the detectors are [not very effective](#). OpenAI [turned off its own AI classifier](#) for text in July 2023 “due to its low rate of accuracy.” The company is currently experimenting with a “provenance classifier”—a tool for identifying images that have been created by Dall-E-3.

If you suspect significant verbatim plagiarism from an AI-generated response, you will almost definitely not be able to find or precisely recreate the “original” source. That said, AI-generated text responses may contain indicative features, particularly when prompts are unrefined and entered by multiple users. These include, but are not limited to, false citations, attributions of scholarly works to authors who did not write them, inaccurate quotations, and generalities or non-committal arguments. Furthermore, though no two responses to the same prompt will be absolutely identical, they may share identical phrases, repetitions, errors, or evidence.

## Institutional and Individual Guidance

Given that inappropriate use is difficult to prove and that not all use is inappropriate, a campus-wide ban at this time would be impracticable and likely counterproductive. AI therefore requires informed and responsible engagement from students and faculty alike.

We recognize that attitudes towards the use of AI may vary divisionally, departmentally, and individually. Faculty will have to decide on a course by course or case by case basis whether or not using AI is ethically or academically acceptable. Any departmental policies

established should be included in your syllabus, as should your individual policies on AI and academic integrity.

We recommend going over them specifically on the first day of class and repeating where appropriate (you may wish to repeat or include them with specific assignments (paper prompts, project descriptions, etc.). You may wish to address the capacities and limitations of AI with students directly and/or collaboratively (a list of activities suggested by the University of Wisconsin may be found [here](#)). Discuss how the technology works; enter a prompt relevant to your course or subject and evaluate the results; explain why using it or not using it matters to you (and should matter to them!) personally, pedagogically, and/or disciplinarily.

Clarity is particularly important because Rhodes students are accustomed to using online tools like Grammarly and may not adequately distinguish between AI-assisted proofreading or editing and the use of AI to generate ideas, content, or structure. One member of the Rhodes Honor Council speculates that students consider it a “gray area;” another suggested that “students recognize that it is cheating but I doubt that they view it as plagiarism or something that extreme.”

In sum, it is incumbent upon faculty to address any ambiguities and remind students that violations of academic integrity, including the unauthorized use of AI or failure to correctly cite it when use is acceptable, may result in penalties for the assignment or course and be reported to the Honor Council.

# Glossary of AI Terms

**Alignment in AI** refers to the idea of designing and training artificial intelligence systems in a way that aligns with human values and goals. It aims to ensure that the AI systems behave ethically, safely, and reliably. In other words, alignment refers to the ability of an AI system to perform tasks that are beneficial to humans while avoiding actions that are harmful or undesirable.

**Artificial Neural Networks (ANN):** A type of machine learning model that mimics the structure and function of the human brain to learn from data and make predictions.

**Bias in AI:** The tendency for AI systems to replicate and amplify societal biases due to the biases in the training data or the algorithms used.

**ChatGPT** is a specific implementation of a type of machine learning called natural language processing (NLP). It is a large language model developed by OpenAI that can generate human-like responses to natural language inputs, making it a useful tool for chatbots and virtual assistants.

**Deep Learning:** A type of machine learning that involves training large neural networks with many layers to learn complex patterns in data.

**Explainable AI:** The ability of an AI system to explain its decision-making process and the factors that influenced its predictions.

**GAN**, "Generative Adversarial Network,": a type of deep learning model used for generative tasks like image or text generation. It consists of two neural networks, a generator and a discriminator, which are trained in a game-like process to improve their performance. GANs have become popular due to their ability to generate high-quality, complex data. **General AI**, also known as Artificial General Intelligence (AGI): a hypothetical artificial intelligence system that can perform any intellectual task that a human can. It is sometimes referred to as "strong AI" and is a long-term goal of AI research.

**Generative AI:** a subset of artificial intelligence that is capable of generating novel and original output that resembles human-created content. It involves the use of complex algorithms and neural networks to create new data or content, such as images, music, and text, without any human intervention.

**GPT-x** refers to a family of generative language models developed by OpenAI that use a deep neural network to generate text. Each model in the GPT series is larger and more complex than the previous one, with the latest version being GPT-3, which is currently one of the largest and most advanced language models in the world.

**Large Language Model:** a type of AI model that can process and generate human-like language. These models are typically trained on large datasets of text and use complex

algorithms to generate natural language outputs. GPT-x is an example of a large language model.

**Machine Learning:** a type of artificial intelligence that involves training a computer program to recognize patterns and learn from data without being explicitly programmed. It is a subset of AI and involves the use of algorithms that can learn from data to make predictions or take actions.

**Natural Language Processing (NLP):** The field of AI that focuses on enabling computers to understand and generate human language.

**OpenAI:** a research organization focused on developing AI. They are known for developing some of the most advanced AI models, such as GPT-4.

**Prompting:** the process of providing a specific input or question to an AI system to generate a desired output. In the context of language models like GPT-x, prompting can be used to generate specific types of text or responses.

**Reinforcement Learning:** A type of machine learning that involves training an AI agent to make decisions in an environment based on a reward system.

**Reinforcement Learning from Human Feedback (RLHF)** is a process whereby designers solicit responses from humans that express their preferences regarding a model's predictions—often in the form of evaluations, ratings, or suggestions. Those preferences are then used to define a more informative and structured reward signal that a reinforcement learning algorithm seeks to maximize. RLHF can enable language models to provide answers that align with these complex values, to generate more verbose responses, and to reject questions that are either inappropriate or outside the knowledge space of the model.

**Text-to-Image:** a type of AI technology that involves generating realistic images from textual descriptions. It uses techniques such as computer vision and generative adversarial networks (GANs) to create images that match the description provided.

**Training:** the process of teaching an AI system to recognize patterns and learn from data by adjusting its internal parameters. In the context of machine learning, training involves providing a dataset to an algorithm and tuning its parameters to optimize its performance.

**Weak AI**, also known as Narrow AI: an AI system that is designed to perform a specific task or a narrow range of tasks. Unlike General AI, Weak AI is limited to specific applications and cannot operate beyond their programmed scope.