Deepfakes

Grade 2-7 Activity Write Up

actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation

DAY OF AI

Deepfakes

Terms of Use

Prior to using this activity or parts thereof, you agree and understand that:

- It is your responsibility to review all aspects of this document and the associated activity write ups, and ensure safety measures are in place for the protection of all involved parties. Any safety precautions contained in the "Safety Considerations" section of the write-ups are not intended as a complete list or to replace your own safety review process.
- Actua shall not be responsible or liable for any damage that may occur due to your use of this content.
- You may adapt the content for your program (remix, transform, and build upon the material), providing appropriate credit to Actua and indicating if changes were made. No sharing of content with third parties without written permission from Actua.
- This material is based on content from Day of AI™ by the Massachusetts Institute of Technology. This and all materials derived from it are subject to the Creative Commons License International 4.0 (CC BY-NC-SA 4.0) and the terms and conditions stipulated by MIT at dayofai.org/privacy-and-terms-of-use ("The License").

About Actua

Actua is creating a Canada where every child has the skills and confidence they need to achieve their full potential. As a leading science, technology, engineering and mathematics (STEM) outreach organization, Actua includes over 40 universities and colleges, engaging 500,000 youth in 600 communities each year. For 25 years, Actua has focused on identifying and removing the barriers for entry into STEM and now have national programs dedicated to engaging Indigenous youth, girls and young women, Black youth, those facing economic barriers and youth in Northern and remote communities. For more information, please visit us online at www.actua.ca and on social media: Instagram, LinkedIn, Facebook and YouTube! For more information, please visit us online at www.actua.ca and on social media: Instagram, Facebook, LinkedIn, TikTok and YouTube!

Deepfakes

Activity Summary

In this activity, participants will learn all about deepfakes, and how to spot them. After being introduced through a video generated by artificial intelligence (AI), participants will learn how this technology uses datasets to create realistic replicas of people, places, and things. They will then act as a Generative AI themselves, combining different features to construct a character. They will then hone their observation skills when they analyze an entirely AI generated advertisement, learning the S.P.O.T. method for thinking critically about media they encounter online.

Developed by Actua, 2025.

Delivery Environment	Activity Duration	Intended Audience	Tech
In-Person	1 Hour	Grades 2-7 (*Ages 6-12*)	**Facilitators should have access to a laptop, projector, speakers, and a screen or blank wall to project onto.** • Projector • Speaker • Screen/Blank Wall • Laptops/Tablets

Achievement Goals

Learning Goals

Following this activity, participants will:

- **Recognize** that Generative AI often draws from real data to generate media, but the final product may not represent reality.
- **Explore** how Generative AI studies specific datasets to create content of specific things/people.
- **Apply** the S.P.O.T. method to evaluate the credibility of online content.

Success Criteria

Following this activity, participants can express:

- **I can explain** that a deepfake is a fake picture, video, or audio clip created by an AI to “look” or “sound” real.
- **I can combine** different elements together to “create” a person, similar to deepfakes.
- **I can ask** critical questions that help me decide whether certain content is real or not.

Logistics (Timing, Group Sizing, Materials)

Section Title	Time	Group Size	Materials
Opening Hook	5 minutes	*Whole Group*	**Facilitators** • Deepfakes 2-7 - Activity Slide Deck (Part 1) *(Appendix C)* • Documentary Clip *(Appendix C)*
Section 1: What are Deepfakes?	10 minutes	*Whole Group*	**Facilitators** • Deepfakes 2-7 - Activity Slide Deck (Part 1) *(Appendix C)*

Section Title	Time	Group Size	Materials
Section 2: The Identity Lab	15 minutes	*Individual*	**Facilitators** • Deepfakes 2-7 - Activity Slide Deck (Part 1) *(Appendix C)* **Per Individual** • Identity Lab Dataset *(Appendix C)* • Writing Utensil • Tape OR Gluestick • Scissors
Section 3: Spot the Bot	20 minutes	*Whole Group*	**Facilitators** • Deepfakes 2-7 - Activity Slide Deck (Part 2) *(Appendix C)* • GeoVitality Commercial (*Appendix C*)
Reflection & Debrief	10 minutes	*Whole Group*	**Facilitators** • Deepfakes 2-7 - Activity Slide Deck (Part 2) *(Appendix C)*

Safety Considerations

Safety considerations have been provided below to support safety during this activity, however they are not necessarily comprehensive. It is important that you review the activity and your delivery environment to determine any additional safety considerations that you should be implementing for the delivery of these activities.

Emotional Safety

- Facilitators should understand that participants have different lived experiences and prior knowledge about AI safety, AI, and digital citizenship. This activity may involve or lead to discussions of sensitive topics, such as ethical implications of AI. Facilitators should encourage open, respectful

discussions and acknowledge all perspectives. Facilitators should always keep the participants' emotional safety in mind in these discussions, and defer to training from their institution and training received.

- This activity involves discussion on misinformation and disinformation, which may be a sensitive topic. Thinking about these issues might remind participants of past experiences or bring up strong feelings. In addition to the suggestions above, facilitators should:
    - Make it clear that participants can step back from any activity, discussion, or resource that feels too heavy.
    - Build in space for reflection, questions, or quiet processing.
    - Offer the following resource on the topic of gendered misinformation and disinformation if relevant, Tackling Online Gendered Disinformation: Youth Guide, which discusses what it is and how to take action. More resources can be found on actua.ca/misinformation.

Cuts

- Remind participants of how to use scissors properly and safely. Ensure participants are sitting down and have age appropriate scissors. Offer assistance when cutting thick materials, such as corrugated cardboard or popsicle sticks to help prevent injury.

Curriculum Links

This activity aligns with these components found in the UNESCO AI Competency Framework for Students:

Human-Centered Mindset: Human Agency

- Learners are expected to be able to recognize that AI is human-led and that the decisions of the AI creators influence how AI systems impact human rights, human-AI interaction, and their own lives and societies (p. 29-30).

Human-Centered Mindset: Human Accountability

- Learners are expected to be able to recognize that human accountabilities are the legal obligations of AI creators and AI service providers, and understand what human accountabilities they should assume during the design and use of AI. They should also foster an awareness that human accountability is a legal and social responsibility when using AI to assist decisions that affect humanity and uphold the principle that humans should not cede the determination to AI when making high-stakes decisions. They are also expected to enhance their judgement on, and attitudinal resilience to, the illusive claims on the use of outputs and as well as predictions that AI can usurp humans' thinking and decision-making (p. 37-38).

Human-Centered Mindset: AI Society Citizenship

- Learners are expected to be able to build critical views on the impact of AI on human societies and expand their human-centred values to promoting the design and use of AI for inclusive and sustainable development. They should be able to solidify their civic values and the sense of social responsibility as a citizen in an AI society. Learners are also expected to be able to reinforce their open-minded attitude and lifelong curiosity about learning and using AI to support self-actualization in the AI era (p. 45-46).

Ethics of AI: Embodied Ethics

- Learners are expected to be able to develop a basic understanding of the ethical issues around AI, and the potential impact of AI on human rights, social justice, inclusion, equity and climate change within their local context and with regard to their personal lives. They will understand, and internalize the following key ethical principles, and will translate these in their reflective practices and uses of AI tools in their lives and learning: Do no harm, Proportionality, Nondiscrimination, Sustainability, Human determination, and Transparency (p. 31-32).

Ethics of AI: Safe and Responsible Use

- Learners are expected to be able to carry out responsible AI practices in compliance with ethical principles and locally applicable regulations. They are expected to be conscious of the risks of disclosing data privacy and take measures to ensure that their data are collected, used, shared, archived and deleted only with their deliberate and informed consent. They are also expected to be conscious of typical AI incidents and the specific risks of certain AI systems, and be able to protect their own safety and that of their peers when using AI (p. 39-41).

AI Techniques and Applications: AI Foundations

- Learners are expected to develop basic knowledge, understanding and skills on AI, particularly with respect to data and algorithms, and understand the importance of the interdisciplinary foundational knowledge required for gradually deepening understanding of data and algorithms. They should also be able to connect conceptual knowledge on AI with their activities in society and daily life, concretizing a human-centred mindset and ethical principles through an understanding of how AI works and how AI interacts with humans (p. 32-34).

AI Systems Design: Problem Scoping

- Learners are expected to be able to understand the importance of 'AI problem scoping' as the starting point for AI innovation. They are also expected to acquire the knowledge and project-planning skills needed in order to conceptualize and construct an AI system (p. 35).

AI Systems Design: Iteration and Feedback

- Learners are expected to enhance and apply their interdisciplinary knowledge and practical methods to evaluate the humanistic appropriateness and methodological robustness of an AI model and its impact on individual users, societies and the environment. They should be able to acquire age-appropriate technical skills to improve the quality of datasets, reconfigure

algorithms and enhance architectures in response to results of tests and feedback. They should be able to apply human-centred mindset and ethical principles in simulating decision-making on when an AI system should be shut down and how its negative impact can be mitigated. They are also expected to cultivate their identities as co-creators in the larger AI community (p. 50-51).

This activity can be connected to the following subject areas:

Science

- Investigating systems with specific inputs, processes, and outputs.

English

- Demonstrate an understanding of how some digital media can be created and how they affect the audience.

Mathematics

- Collecting, organizing, and interpreting qualitative and quantitative data.

Community Connections

Community connections are suggestions from Actua, grounded in our approach, on how facilitators can adapt the activity to reflect the strengths, interests, and priorities of the community where or with whom it is delivered. Consider the following guiding questions to adapt the activity in meaningful ways:

- **Consult with community:** Are there local organizations, Knowledge Keepers, or community members who could contribute insight or context to this topic?
- **Draw on youth experience:** How can you give participants opportunities to share, reflect on, and apply how this learning is relevant to them or their community? Invite participants to identify what knowledge, who, and where they already learn from.

- **Integrate local examples:** How can you tailor this activity to local or regional interests, industries, or community priorities (e.g. land and environment, health, technologies)?

Activity Procedure

To Do in Advance

SECTION	PREPARATION

General	• **Think ahead and be ready to adapt:**
	  ◦ Determine your **delivery method** and leverage ideas from the delivery recommendations and adaptations sections.
	  ◦ While **estimated times** are provided, it will be helpful to think about how much time you would like to spend on different activities and discussions.
	  ◦ While **group sizes** (individual, pairs, groups) are suggested, many activities are flexible for whatever will work in your classroom.
	• **Prepare for the content:**
	  ◦ Have **answers in mind** to share with participants for the various reflection questions asked.
	  ◦ Examine the provided materials to determine if they are **suitable** for your participants.
	• **Equipment:**
	  ◦ Ensure device, screen and projector are set up.
	  ◦ Prepare participant devices.

Section 2: The Identity Lab	• Prepare printed copies of the Identity Lab Dataset *(Appendix C)*.

Opening Hook

1. Play the opening “Documentary Clip” (*Appendix C*) for participants, telling them that you have something to show them from a recent nature documentary that has gone viral!
    a. As they watch the clip, facilitators should act impressed and genuinely excited, as though they are taking the clip completely seriously.
2. After the clip finished, ask participants what they think!
    a. There will likely be some pushback - some may even say its AI generated! If they do, ask about their reasoning - be sure not to say they are wrong.
3. After a bit of back and forth, come clean - You know this clip isn’t real! But there is a lot of media out there like this that is tricking people into believing things that aren’t true.
4. “Does anyone know what this kind of media is called?”
    a. Deepfakes!

Section 1: What are Deepfakes?

1. Using the Deepfakes 2-7 - Activity Slide Deck (Part 1) (*Appendix C*), ask participants if anyone is familiar with AI, generative AI or deepfakes? Provide some time for them to share their thoughts.
    a. The most “mainstream” use of the technology associated with Deepfakes is the process of de-aging actors in film and TV. One of the most popular examples of this is Luke Skywalker in The Mandalorian / Book of Boba Fett - In both of these, a different actor played the role, and then their face was swapped out for a younger version of Mark Hamill.
2. Explain to participants the definition of each:
    a. **Artificial Intelligence (or AI)** refers to algorithms created by people that help machines act smart, they can learn, solve problems, and make decisions. These algorithms teach computers to use the data they have,

like pictures, sounds, or repeated experiences, to recognize patterns, make predictions, and improve over time.

   b. **Generative AI (GenAI)** is the broad technology of AI models that create new content (like text, images, or video) from large datasets, while deepfakes are specific applications of generative AI used to create realistic but fake content that mimics a real person's likeness, voice, or actions. Sometimes it is meant to be harmful.
   c. **Deepfakes** are content that is generated or manipulated using GenAI with the intention of passing it off as real. This content is usually images or video, but can be audio and text as well (or any combination of these things).
      i. Put simply - a deepfake is digital media that is made or modified using AI that is then presented as “real”.

3. Deepfakes are often “copies” of specific things, like people, animals, or objects. It then either places the copy it created in a real scene (like adding a car into a pre-existing image), or creates the entire scene from scratch (like in a picture of someone who doesn’t exist).
4. Remind participants that Generative AI requires **datasets** to create.
   a. Datasets function like an AI’s “memory”. They are collections of datapoints (these can be images, videos, text, and more) that function as a reference for the AI to study over and over again. They are very broad in the data they contain for general purpose Gen AI, but deepfakes require more specificity.
5. To create a deepfake of something specific, like a face, it needs a dataset entirely based on an individual's face. This takes the form of images of the person. Videos can be used as well, but it will extract every frame from the video to study, so it functions like images!
   a. For de-aging in film and TV, the dataset comes from movies and shows that the actor was in when they were younger!
6. The AI will then study this dataset many times, learning the face’s features from all different angles, lighting and environments. Once it has done so, it

can either create an original image of the face from scratch, or modify a pre-existing image by “placing” a copy of the face onto another person's body.

Section 2: The Identity Lab

1. Explain to participants that deepfakes don’t have to be of a specific person. Generative AI can also use datasets of many different people and combine them to create deepfakes of people who don’t exist.
2. Inform participants that they will be acting as an AI creating a deepfake of a “new” person!
3. Distribute the Identity Lab Dataset (*Appendix C*) to participants, explaining that this will act as the dataset for their creation. It has different face shapes, eyes, mouths, hairstyles, and hands! Their task as the AI is to choose different features and combine them to create a “new” person.
4. Provide scissors, a glue stick or tape, and a writing utensil to each participant.
5. Participants can either cut features out and tape them onto the face of their choosing (example on the right), or they can trace them by placing the face shape of their choice on top of the features they want and outlining them (example on the left).
6. Provide participants time to create their characters. Facilitators should circulate around to offer support.

Section 3: SPOT the Bot

1. Using the Deepfakes 2-7 - Activity Slide Deck (Part 2) (*Appendix C*), Remind participants that deepfakes create images of people, animals, objects, and places that are **not** real. Sometimes they are completely generated, meaning the content doesn’t actually exist in the real world - just like the characters you made!

2. Deepfakes are often used to create content that replicates a real individual in situations they have never been in. Because of this, it is very important to learn about possible signs of deepfakes!
    a. When deepfakes are used in this context, they are spreading **mis/disinformation**.
        i. **Misinformation** is untrue content that is spread by people who believe that it is true. Misinformation could be spread innocently, or to cause harm.
        ii. **Disinformation** is untrue content that is spread by people who know that it is untrue. Disinformation is always spread knowingly and deliberately to cause harm
3. Inform participants that they are going to watch a video, and need to try and guess what things were generated by AI.
    a. Show the GeoVitality Commercial (*Appendix C*).
4. Ask participants: "What do you think was AI Generated?"
    a. After some guesses, reveal the answer: It was ALL generated by AI!
5. Go through the slides and discuss different aspects of the commercial, how it was made, and ways to potentially notice/spot that it is AI.
    a. **Video:**
        i. All of the video content was generated by VEO 3, Google's AI video generator. We put the ideas into Gemini, Google's text generator, and asked it to create prompts for VEO!
        ii. Spotting the AI video: Highlight some of the messed up hands, slightly blurry footage, uncanny faces.
    b. **Audio:**
        i. Both the background song and the voiceover were generated - The song was generated by Suno, and the voice was generated by ElevenLabs.
        ii. AI audio can be very tricky to spot! The music is almost impossible to notice. For the voiceover, the only hint is that there is some slightly unnatural sounding speech in the beginning, but even that is minimal!

    c. **Images:**
        i. In this video, there was only one image - the logo! This was generated by Google Gemini.
        ii. In this case, there is no way to really spot or notice that it is AI generated. But that is not always the case for images, like we discussed before with people!
6. Whenever you are reading, watching, or listening to anything on the internet, it is important to think critically and ask yourself: “Does this make sense?” or “Does this seem like something that person would say or do?”.
7. If you are ever unsure about whether or not content is a deepfake or AI generated, use a technique called S.P.O.T!
    a. **S**: Is this a credible **S**ource?
        i. Where does it come from? Is it a random social media account, or shared on the news? Have you found any other reliable content from this source before?
    b. **P**: Is the **P**erspective biased?
        i. Does the person sharing this have a reason they would want you to believe whatever it is?
    c. **O**: Are **O**ther sources reporting the same story?
        i. Did you recognize this photo, video, or information anywhere else? If you research it, is anyone else discussing it?
    d. **T**: Is the story **T**imely?
        i. When was it posted? Are they referencing old information, or claiming to be referencing a brand new study?
8. To decide whether or not we would want to use S.P.O.T on the video we just watched, let’s ask the important question: “Does it make sense?”
    a. No! Eating rocks does not make sense.
9. Let’s use S.P.O.T on the video we just watched!
    a. **S:** In the video, it says that eating rock is backed by “research”, but it doesn’t reference anything other than that! There is also someone dressed up as a doctor in the video, but there is no name or credits to who this “doctor” is!

b. **P:** The video is an advertisement for a company called GeoVitality, which means it is biased to make them look good! If they are trying to sell this, of course it will be a positive video.
c. **O:** If you do some googling on if eating rocks is good for you, something interesting appears - an incident that happened about a year ago where Google's AI Overview said that eating a rock or two a day is healthy. This was quickly disproven and Google made sure to fix that part of their AI!
d. **T:** This video is brand new! That means it was created during a time where AI content is very easy to create, and doesn't credit anyone or any research at all.

Reflection & Debrief

1. Gather participants for a group discussion to reflect on what they learned.
2. Ask participants: "What are deepfakes?"
    a. Deepfakes are images, videos, and/or audio generated or manipulated by AI with the intention of passing it off as real.
2. Ask participants: "How could videos like the GeoVitality commercial potentially be harmful?"
    a. **Mis/Disinformation!** This video, and others like it, promote false information as though it is fact.
3. Ask participants: "What are some clues to figure out if something is a deepfake?"
    a. Visual artifacts - is there anything weird with the hands, faces, backgrounds, etc? Does it seem uncanny?
    b. Does it sound natural, or kind of like a robot?
    c. **Does it make sense?**
4. Ask participants: "If you aren't sure whether or not something is potentially a deepfake, what can you use to help figure it out?"
    a. **S.P.O.T!**
5. Discuss the different careers listed in *Appendix A: Career & Mentor Connections*.

Delivery Adaptations

Modifications are ways to make the activity more accessible, **extensions** are ways to make the activity last longer or more challenging.

Modifications

SECTION 2: THE IDENTITY LAB

- Facilitators can work with participants as a group to create some characters from the dataset, taking votes on which features to use for the character.

SECTION 3: SPOT THE BOT

- For younger participants, S.P.O.T can be simplified into three questions:
    - Does this make sense?
    - Who made this?
    - Why did they make it?

Extensions

SECTION 2: THE IDENTITY LAB

- Introduce simplified prompts to the character creation process. Facilitators can create several prompts for participants to choose from, or they may come up with their own and share it at the end. (For example, “Create a character that looks like a genius!” or “Create a character who looks like someone you know!”).
- Once participants have created their characters, they can play a guessing game that emulates the generation/validation process of a Generative Adversarial Network (the model most commonly used to create deepfakes). Put participants in pairs, hiding their characters from each other. They will then ask simple yes/no questions about the other participants' character (Do they have glasses, do they have short hair, etc) and attempt to reconstruct their partner's character!

ADDITIONAL RESOURCE

- If participants are interested in learning more about gendered misinformation and disinformation, the following resource can be shared: Tackling Online Gendered Disinformation: Youth Guide. Refer to the Safety Considerations section for relevant considerations that would apply here.

References & Gratitude

Actua. (2025). *Misinformation*. actua.ca/misinformation

Amazon Web Services. (n.d.). *What is Machine Learning?*. Amazon. aws.amazon.com/what-is/machine-learning/

Badman, A. & Kosinski, M. (n.d.). *What is Data?*. IBM. ibm.com/think/topics/data

Bartels, M. (2024). *How to Tell If a Photo Is an AI-Generated Fake.* Scientific American. scientificamerican.com/article/how-to-tell-if-a-photo-is-an-ai-generated-fake/

BBC Newsround. (2024). *Deepfake technology: What is it, how does it work, and what can it be used for?* bbc.co.uk/newsround/69009887

Davenport, T., Mittal, N. (2022) *How Generative AI Is Changing Creative Work*. Harvard Business Review. hbr.org/2022/11/how-generative-ai-is-changing-creative-work

Day of AI. (2025). *How Do Machines Create?*. dayofai.org/units/how-do-machines-create

Generative AI – What is it and How Does it Work? (n.d.). NVIDIA. nvidia.com/en-us/glossary/generative-ai/

Kamali, N., Nakamura, K., Chatzimparmpas, A., Hullman, J., & Groh, M. (2024). *How to Distinguish AI-Generated Images from Authentic Photographs*. arxiv.org/abs/2406.08651

Khanjani, Z., Watson, G., & Janeja, V. P. (2023). Audio deepfakes: A survey. doi.org/10.3389/fdata.2022.1001063

Lawton, G. (2024). *What is Gen AI? Generative AI explained*. Search Enterprise AI. techtarget.com/searchenterpriseai/definition/generative-AI

Morrison, S. (2023). *The tricky truth about how generative AI uses your data*. Vox. vox.com/technology/2023/7/27/23808499/ai-openai-google-meta-data-privacy-nope

Proofpoint. (2025). *What Is a Deepfake? Definition & Technology* proofpoint.com/us/threat-reference/deepfake

The Alan Turing Institute. (2024). *What are deepfakes and how can we detect them?* turing.ac.uk/blog/what-are-deepfakes-and-how-can-we-detect-them

United Nations Educational, Scientific and Cultural Organization. (2024). *UNESCO AI framework: Competency levels*. unesdoc.unesco.org/ark:/48223/pf0000391105

University of Toronto Libraries. (2025). *Artificial Intelligence for Image Research: How Generative AI Models Work*. guides.library.utoronto.ca/c.php?g=735513&p=5297039

U.S. GAO. (2020). *Deconstructing Deepfakes—How do they work and what are the risks?* gao.gov/blog/deconstructing-deepfakes-how-do-they-work-and-what-are-risks

Zewe, A. (2023). *Explained: Generative AI.* Massachusetts Institute of Technology. news.mit.edu/2023/explained-generative-ai-1109

Appendices

Appendix A: Career & Mentor Connections

AI/MACHINE LEARNING ENGINEER

- Builds and trains artificial intelligence systems that can learn from data to make predictions or perform complex actions. They support the machine learning researcher.

ARTIST

- Artists work in a variety of mediums to create original, creative work. An artist can be created for museums, advertising, businesses or a variety of other purposes, such as personal expression.

COMPUTER PROGRAMMER

- A computer programmer is a person who creates computer software. They write code to build websites, computer games, financial analysis and many more.

MACHINE LEARNING RESEARCHER / DATA SCIENTIST

- Machine learning researchers or data scientists clean and interpret data while building models using a combination of that data and machine learning algorithms.

STORYTELLER

- Storytellers are experts at telling stories using a variety of mediums which can include performing, writing, consulting, and podcasting.

Appendix B: Background Information

DATA AND DATASETS

Data comes in many forms, including numbers, images, words, and other types of information. They provide insights about individuals, groups, and the world around us. It is a key resource for understanding patterns, making decisions, and predicting outcomes. Specifically, data can be used to:

- Identify correlations and trends that might not be obvious at first glance.
- Develop a deeper understanding of complex systems and behaviors.
- Inform decisions in real time or future planning.
- Make decisions based on past patterns or observed information.

Datasets are also essentials for creating algorithms, which are sets of instructions or rules that a computer follows to solve problems or make decisions. By analyzing and learning from data, algorithms can help humans understand and act on information more effectively.

Some applications of how data drives algorithms include:

- **Traffic and navigation apps:** Analyzing road and traffic patterns to suggest the fastest routes.
- **Outer space and weather systems**: Studying patterns to predict astronomical events or weather changes.
- **Shopping and media recommendations**: Using purchase and viewing history to suggest products or content that people may like.
- **Healthcare and diagnostics**: Analyzing medical data to identify health risks, suggest treatments, or support research.

ARTIFICIAL INTELLIGENCE

Artificial intelligence (AI) is a branch of Computer Science that deals with a machine's ability to simulate intelligent behaviour. This includes cognitive functions we associate with human minds, such as perceiving, reasoning, learning, and adapting.

AI is becoming increasingly vital in our lives. From digital assistants, GPS navigation, and autonomous vehicles to tools like Siri/Google Home and generative AI tools (e.g., OpenAI's Chat GPT), its impact on our daily lives is growing. AI plays a crucial role in various aspects of work, enhancing efficiency, and taking on hazardous or monotonous tasks. As AI applications grow, discussions on AI ethics and responsible practices are increasingly important.

MACHINE LEARNING

Machine learning (ML) is a type of artificial intelligence where computers learn from data, improve at tasks over time, and adapt without being programmed step by step. Instead of giving the computer every possible instruction, we give it ways to learn from experience, similar to how people practice and improve.

For example, you might teach a computer to play checkers. Rather than listing every possible move, you could program it to learn from playing many games. Eventually, it could get so good that it plays better than the person who wrote the program

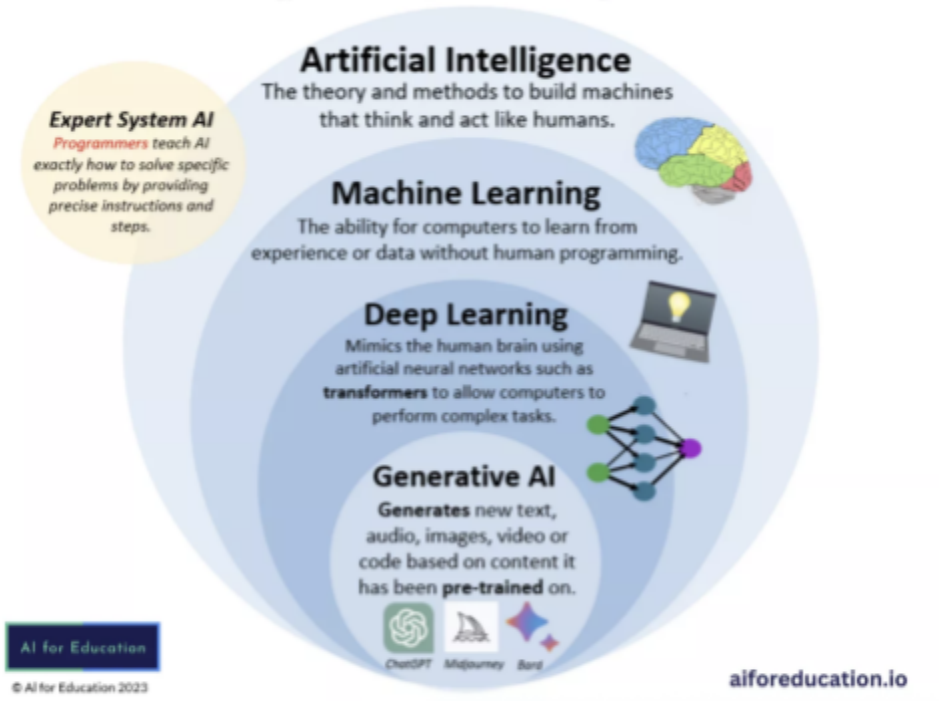
The learning process usually includes a few steps:

1. **Give it Data:** First, the computer is given a lot of data to process. This could be pictures, numbers, text, or sounds.
2. **Find the Patterns:** The computer then analyzes all this data and starts to find patterns, building its own “rules” based on experience.
3. **Make Predictions or Actions:** Using these patterns, the computer can make decisions or predictions when it encounters new data.
4. **Get Better and Better:** As the computer gets more data, it can adjust its patterns to become more accurate. The more data and computing power it has, the better its predictions become

Machine learning is used across different fields and in our everyday lives. In your daily life, ML works quietly behind the scenes in many of the tools and apps you use. It powers autocorrect and predictive text on your phone, suggests videos to watch on streaming platforms, and runs the filters on social media apps that recognize and

track your face. It also helps your phone unlock through facial recognition by learning your unique features, and it keeps your email inbox clean by filtering out junk mail.

In addition, ML is making an impact in the wider world. In healthcare, it helps doctors diagnose illnesses more accurately, such as spotting broken bones in X-rays. It is also the “brain” behind self-driving cars, enabling them to observe and respond to the world around them. In industries like agriculture and manufacturing, ML supports smarter ways to grow food and produce goods more efficiently.



AI for Education. (2023). *Generative AI Explainer*. aiforeducation.io/ai-resources/generative-ai-explainer

Artificial Intelligence VS Machine Learning VS Deep Learning

You might come across the terms “artificial intelligence” and “machine learning” used together, but they aren’t exactly the same thing. AI is the broad idea of creating machines that can act or think in ways similar to humans. This can range from simple tools like a smart thermostat that adjusts the temperature based on rules, to more advanced systems like voice assistants or even robots. AI systems can learn over time, sense their environment, and make decisions on their own. Within AI, there are different types:

- **Classification AI**: Systems that identifies and sorts things (e.g. Tiktok or Snapchat face filters).
- **Predictive AI:** Systems that use data to make decisions about the future (e.g. Spotify and Apple Music recommendations).
- **Generative AI:** Systems which create new things, such as text, images, or music (e.g. CoPilot and ChatGPT).

Machine Learning is a specific approach within AI. It's a specific way of achieving AI by having computers learn from data instead of being programmed with a long list of rules. So, all machine learning is AI, but not all AI uses machine learning.

Deep Learning (DL) is a more advanced type of ML that uses structures called neural networks, designed to work a bit like the human brain. Deep learning is what enables high-performance AI systems, including many generative AI tools that can create new and transformative outputs.

Types of Machine Learning

Machine learning methods are usually grouped into three main types, depending on the goal and the kind of data they use.

- **Supervised learning** happens when the computer is trained with examples that already have the right answers (called labels). The computer learns the link and patterns between the input and output so it can predict answers for new data. This is used in things like email filters that sort *spam vs. not spam*, or in predicting numbers, like house prices.
- **Unsupervised learning** is when the computer is given data without any labels and has to find patterns on its own. It might group similar things together (clustering), or uncover hidden connections in the data. This is used for things like grouping customers into types or making product recommendations.
- **Reinforcement learning** is like learning by trial and error. A computer "agent" makes decisions in an environment, gets rewards or penalties for its actions, and learns strategies to do better over time. This approach is often used in robotics and training AI to play video games.

GENERATIVE AI

Generative AI is a type of artificial intelligence (AI) designed to create new content, such as text, images, music, or code, by learning patterns from existing data. As a subset of AI, which broadly refers to machines performing tasks that normally require human intelligence, generative AI specifically focuses on producing original outputs rather than just analyzing or recognizing information.

Generative AI models generate new data that resembles the examples they were trained on by understanding underlying patterns and structures. Instead of simply responding to inputs with predefined answers, generative AI can create novel and creative content.

What can generative AI create?

- **Text:** Stories, poems, essays, summaries, chat responses, reports, and even programming code.
- **Images:** Drawings, paintings, photorealistic pictures, designs, and digital art from descriptions or sketches.
- **Audio:** Music compositions, sound effects, voice synthesis, and speech generation.
- **Video:** Short animations, deepfake videos, or video sequences from textual prompts (in emerging applications).
- **3D Models:** Shapes and objects for games, simulations, or design prototypes.

Examples of Generative AI:

- **ChatGPT and other large language models:** Generate human-like text based on prompts.
- **DALL·E and Midjourney:** Create images from textual descriptions.
- **Music generation models:** Compose new songs or melodies.
- **Code generation tools:** Produce programming code based on natural language instructions.

DEEPFAKES

Deepfakes are a category of media where content is generated or manipulated using generative AI with the intent of being passed off as authentic. This content can include images, video, audio, and text. In simple terms, a deepfake is any digital media created or modified by AI that is presented as "real."

While the term is generally associated with manipulating a person's likeness, the definition used in this activity is broader. Deepfakes often create "replicas" of specific subjects, which can include **people, animals, or objects**. The AI can then place this replica into a real scene (e.g., adding a photorealistic, AI-generated car into an existing photograph) or generate the entire scene from scratch (e.g., creating a portrait of a person who does not exist).

This process relies on training an AI model on a highly specific **dataset**. For general generative AI, datasets are very broad. For creating a convincing deepfake “replica”, the dataset must be more focused. For example, to replicate a specific object, the AI would study thousands of images of that object from multiple angles and in various lighting conditions until it can generate a new, convincing version.

How are Deepfakes Created?

Visual deepfakes are most commonly created using a **Generative Adversarial Network (GAN)**. A GAN consists of two distinct neural networks that are trained simultaneously in a competitive, or "adversarial," process:

1. **The Generator:** This network's function is to create the content. It attempts to generate an image that accurately mimics the features found in its training dataset.
2. **The Validator (or Discriminator):** This network acts as an expert critic. It is trained on the dataset of the target subject (e.g., real photos of the object or person being replicated) and its job is to determine if the images from the Generator are authentic or fake.

The process is an iterative feedback loop. The Generator creates images, and the Validator judges them. Based on this feedback, the Generator adjusts its approach to get better at fooling the Validator, while the Validator gets better at spotting fakes. This cycle is repeated many (upwards of millions) times until the Generator can produce highly convincing replicas that the Validator can no longer easily distinguish from real examples.

Spotting Deepfakes

Before analyzing technical details, the easiest way to start is by asking some key questions about the content's context and intent:

- **Does this make sense?**
    - Does the situation seem too good to be true, strange, or "uncanny"? If the content features a known person, does their behavior or speech align with what is publicly known about them? Content that feels out of character or context is a primary red flag.
- **Does this want something from me?**
    - Is it aggressively trying to persuade you to purchase a product, sign up for a service, or adopt a specific viewpoint? Content with a strong persuasive or urgent call to action warrants higher scrutiny.
- **Does this have signs of being AI-generated?**
    - This involves looking for specific technical flaws or artifacts left behind by the AI generation process. See the next section for more information on artifacts to look for.

Identifying Technical Artifacts

- **Video and Images:**
    - **Unnatural Faces:** Look for an "uncanny valley" effect where faces seem slightly off. Blinking may be infrequent or unnatural, and facial expressions may not match the emotional tone of the situation.
    - **Hands and Hair:** AI models frequently struggle to generate hands correctly, resulting in an incorrect number of fingers, strange

proportions, or unnatural bending. Hair may also appear unrealistic, with individual strands looking blurry, stringy, or blending unnaturally into the background.

- **Inconsistent Details:** Check for mismatched earrings, asymmetrical details on clothing, or reflections in glasses or eyes that do not match the surrounding environment.
- **Blurry or Warped Backgrounds:** Pay attention to the area where the subject meets the background. This area may be blurry, warped, or contain nonsensical shapes as the AI struggles to blend the foreground and background seamlessly.
- **Lighting and Shadows:** Inconsistent lighting on the subject compared to the environment, or shadows that fall in the wrong direction, are often indicators of manipulation.

- **Audio:**
  - **Unnatural Cadence:** AI-generated voices can sometimes have a flat, monotonous tone, lacking the natural emotional inflection of human speech.
  - **Strange Pacing or Pauses:** Listen for awkward pacing, unnatural pauses between words, or robotic-sounding pronunciation.
  - **Lack of Background Noise:** Authentic recordings usually contain subtle background noises or the sound of the speaker taking a breath. The absence of these can be a sign of a sterile, AI-generated audio clip.

The S.P.O.T. Method

The **S.P.O.T.** method provides a structured framework for verifying the source and credibility of the content itself.

- **S:** Is this a credible **Source**?
  - Where does it come from? Is it a random social media account, or shared on the news? Have you come across any other reliable content from this source before?

- **P:** Is the **Perspective** biased?
    - Does the person sharing this have a reason they would want you to believe whatever it is?
- **O:** Are **Other** sources reporting the same story?
    - Do you recognize this photo, video, or information anywhere else? If you research it, is anyone else discussing it?
- **T:** Is the story **Timely**?
    - When was it posted? Are they referencing old information, or claiming to be referencing a brand new study?

ETHICS AND AI

Artificial intelligence offers powerful tools and new possibilities. As these systems learn from data, make decisions, and shape our world, it is important to consider their ethical impacts.

Actua has developed a resource (*Appendix C*) to support facilitators in leading discussions with youth about ethics and responsible AI use. Facilitators are encouraged to engage youth in meaningful conversations that empower them to think critically about how AI is designed, used, and experienced in the world around them. This resource emphasizes human agency and responsibility, supports values-based reflection, and creates space for curiosity, dialogue, and informed decision-making as digital citizens.

Appendix C: Additional Resources

GENERAL

Activity Slide Deck

- Deepfakes 2-7 - Activity Slide Deck (Part 1)
    - ***Note:*** This link will automatically download to your device.
- Deepfakes 2-7 - Activity Slide Deck (Part 2)
    - ***Note:*** This link will automatically download to your device.

Supporting Resource

- AI in Context: Responsibility and Ethics in Artificial Intelligence

OPENING HOOK

Video(s)

- Documentary Clip

SECTION 2: THE IDENTITY LAB

Activity Page(s)

- Identity Lab Dataset (refer below)

SECTION 3: SPOT THE BOT

Video(s)

- GeoVitality Commercial

Deep Fakes

Identity Lab Dataset

actüa
Youth · STEM · Innovation
Jeunesse · STIM · Innovation