Glossary

Terms and definitions from the course

A

Artificial intelligence (AI): Computer programs that can complete cognitive tasks typically associated with human intelligence

AI augmentation: The process of using AI to improve a work product, whether by making it easier to do or higher in quality

AI automation: The process of using AI to accomplish tasks, without any action on the user's part

AI model: A computer program trained on a set of data to recognize patterns and perform specific tasks

AI tool: AI-powered software that can automate or assist users with a variety of tasks

AI user: Someone who leverages AI to complete a personal or professional task

Allocative harm: Wrongdoing that occurs when an AI system's use or behavior withholds opportunities, resources, or information in domains that affect a person's well-being

B

Biased data: Data that is incomplete, does not accurately represent populations, or includes preferential treatment for certain individuals or groups

C

Chain-of-thought prompting: A prompting technique that involves requesting a large language model to explain its reasoning processes

Cognitive task: Any mental activity, such as thinking, understanding, learning, and remembering

Conversational AI tool: A generative AI tool that processes text requests and generates text responses

D

Data bias: A circumstance in which systemic errors or prejudices lead to unfair or inaccurate information, resulting in biased outputs

Deepfakes: AI-generated fake photos or videos of real people saying or doing things that they did not do

Drift: The decline in an AI model's accuracy in predictions due to changes over time that are not reflected in the training data

F

Few-shot prompting: A technique that provides two or more examples in a prompt

G

Generative AI: AI that can generate new content, like text, images, or other media

H

Hallucinations: AI outputs that are not true

Human-in-the-loop approach: A combination of machine and human intelligence to train, use, verify, and refine AI models

I

Interpersonal harm: The use of technology to create a disadvantage to certain people that negatively affects their relationships with others or causes a loss of one's sense of self and agency

K

Knowledge cutoff: The concept that an AI model is trained at a specific point in time, so it doesn't have any knowledge of events or information after that date

L

Large language model (LLM): An AI model that is trained on large amounts of text to identify patterns between words, concepts, and phrases so that it can generate responses to prompts

M

Machine learning (ML): A subset of AI focused on developing computer programs that can analyze data to make decisions or predictions

Multimodal model: An AI model that can accept and learn from multiple types of input, such as images, video, or audio

N

Natural language: The way people talk or write when communicating with each other

O

One-shot prompting: A technique that provides a single example in a prompt

Open dataset: A dataset that is freely available to anyone to use

P

Privacy: The right for a user to have control over how their personal information and data are collected, stored, and used

Prompt: Text input that provides instructions to the AI model on how to generate output

Prompt engineering: The practice of developing effective prompts that elicit useful output from generative AI

Q

Quality-of-service harm: A circumstance in which AI tools do not perform as well for certain groups of people based on their identity

R

Representational harm: An AI tool's reinforcement of the subordination of social groups based on their identities

Responsible AI: The principle of developing and using AI ethically, with the intent of benefiting people and society while avoiding harm

S

Security: The act of safeguarding personal information and private data, and ensuring that the system is secure by preventing unauthorized access

Social system harm: Macro-level societal effects that amplify existing class, power, or privilege disparities, or cause physical harm, as a result of the development or use of AI tools

Systemic bias: A tendency upheld by institutions that favors or disadvantages certain outcomes or groups

T

Training set: A collection of data used to teach AI

Transparency: The idea that an AI tool should provide insight into how it works, why it made a particular output, and what factors contributed to that output