



Refining AI-Generated and Ethical Practices

– In collaboration with NASA GRC

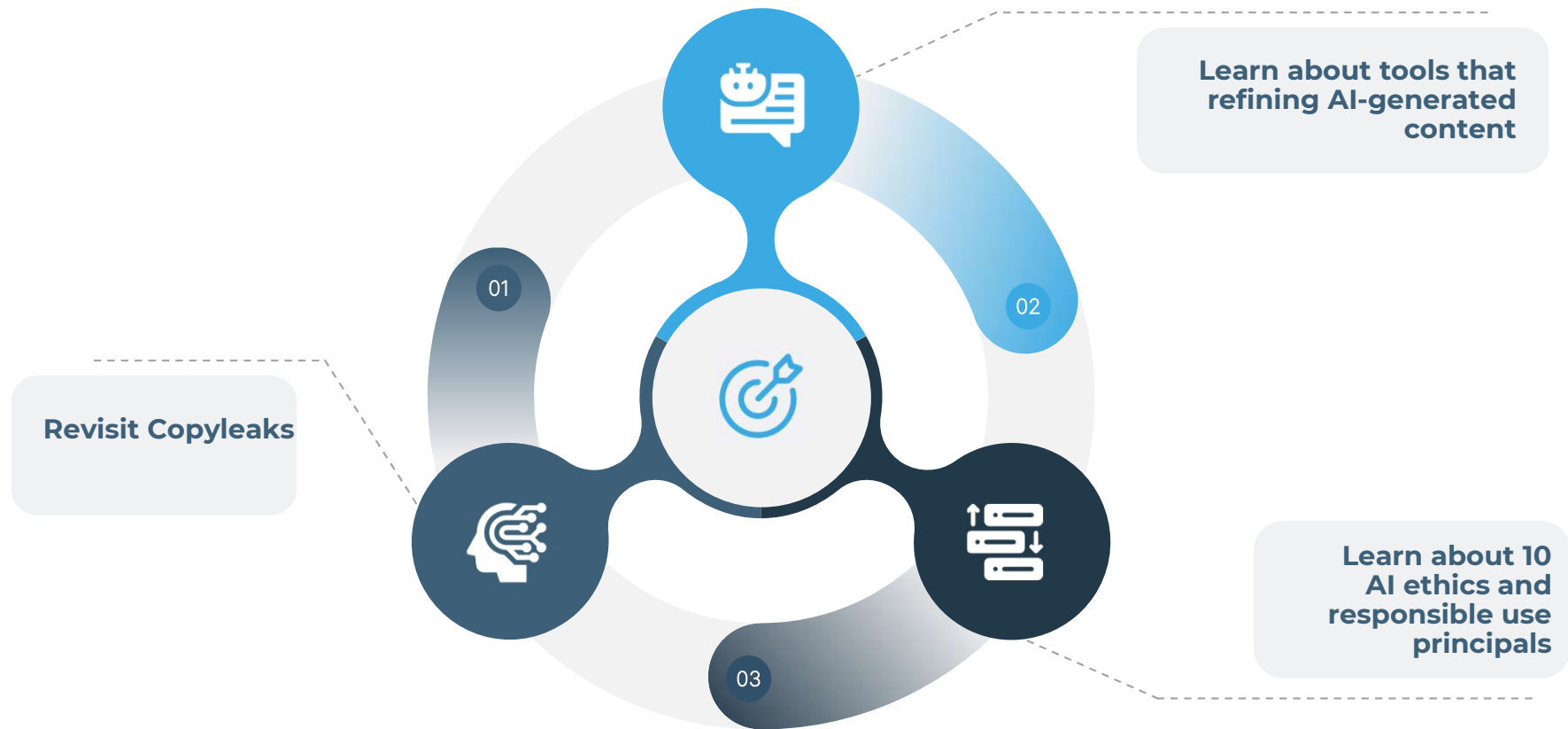
DECEMBER 11, 2024

Agenda



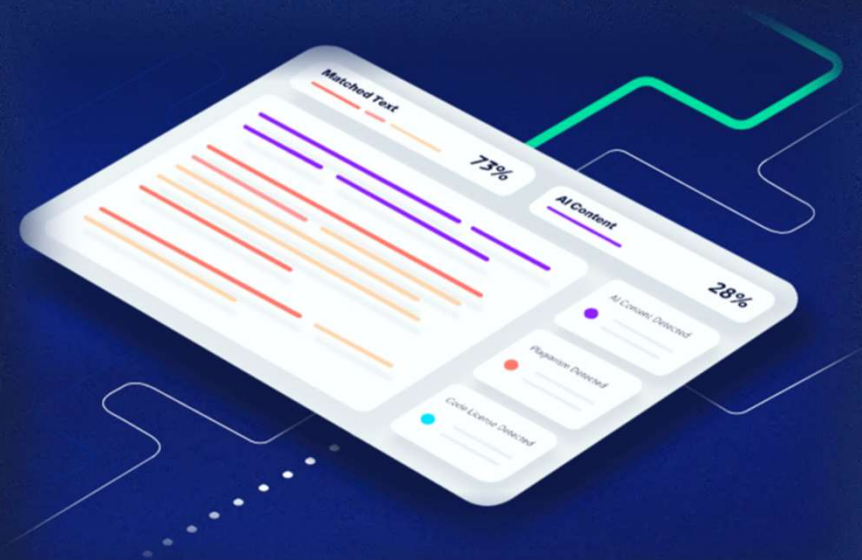
Breakfast and Icebreaker	9:00 – 9:30
Revisit: Copyleaks Learn and Lab: Code only	9:30 – 10:00
Refining AI-Generated Content Learn and Lab	10:00 – 10:35
Learn: Ethical Practices	10:35 – 11:05
Break	11:05 – 11:20
Kahoot round 1	11:20 – 11:40
Revisit Modules 1 and 2	11:40– 11:50
Forward-looking Benefits	11:50 – 12:00
Survey and Quiz	12:00 – 12:20
Wrap-up	12:20 – 12:30

Three take-aways from this training





Copyleaks Learn and Lab



LEARN

Copyleaks

Launched in 2015, Copyleaks began as a plagiarism detection tool but quickly evolved to address the growing need to detect AI-generated content as its prevalence increased.



CAPABILITIES



Text and Code Detection:

Specializes in identifying both written content and programming scripts generated by AI, making it a go-to tool for technical fields.



Multi-Language Support:

Offers detection capabilities in multiple languages, enhancing its usability across global academic and professional contexts.

STRENGTHS



All-in-One Solution:

Combines plagiarism detection and AI content detection in a single platform.



High Accuracy for Code:

Particularly strong in detecting AI-written programming scripts, a feature that sets it apart from many other tools.



Custom Workflow Integration:

Supports API functionality, allowing organizations to incorporate it into their unique workflows.

WEAKNESSES

Limited Multimedia Detection

While Copyleaks excels at detecting text and code, it does not support the detection of AI-generated multimedia content, such as images, audio, or video.

Moderate User Interface Complexity

The interface, while feature-rich, may feel overwhelming for non-technical users or those unfamiliar with advanced detection tools.

Cost

Copyleaks can be expensive for smaller organizations or individual users, especially when requiring advanced features like API integration or multi-language support.

Dependence on Predefined AI Models

Its detection capabilities are limited to the specific AI models it has been trained to recognize, which could lead to challenges as new models emerge rapidly.

False Positives/Negatives

Like other AI detection tools, Copyleaks may occasionally misidentify content, either flagging human-written text as AI-generated (false positive) or failing to detect AI involvement (false negative).

LAB: USE CASE #1 “HUMAN-GENERATED VS AI-GENERATED CODE”

Step 1

Sign Up or Log In

Visit the Copyleaks website and create an account or log in if you already have one.

Step 2

Upload Content for Scanning

- Once logged in, click on "New Scan" from the dashboard.
- Upload the content you wish to scan (e.g., a document, text, or programming code) from your computer. You can also paste the content directly.

Step 3

Select Detection Options

- Choose "AI Detection" for text or "Code AI Detection" for programming scripts.
- For text, you can choose to scan for both plagiarism and AI-generated content.

Step 4

Analyze Results

- After uploading, click "Start Scan" and wait for the results to process.
- Once the scan is complete, Copyleaks will provide a detailed report, highlighting flagged sections and indicating the percentage of AI-generated content.
- Review the report to identify sections of the content that may have been written by AI.

Step 5

Export Results

You can download the report in PDF format or share it with colleagues using the provided sharing options.

Step 6

Hands-On Practice

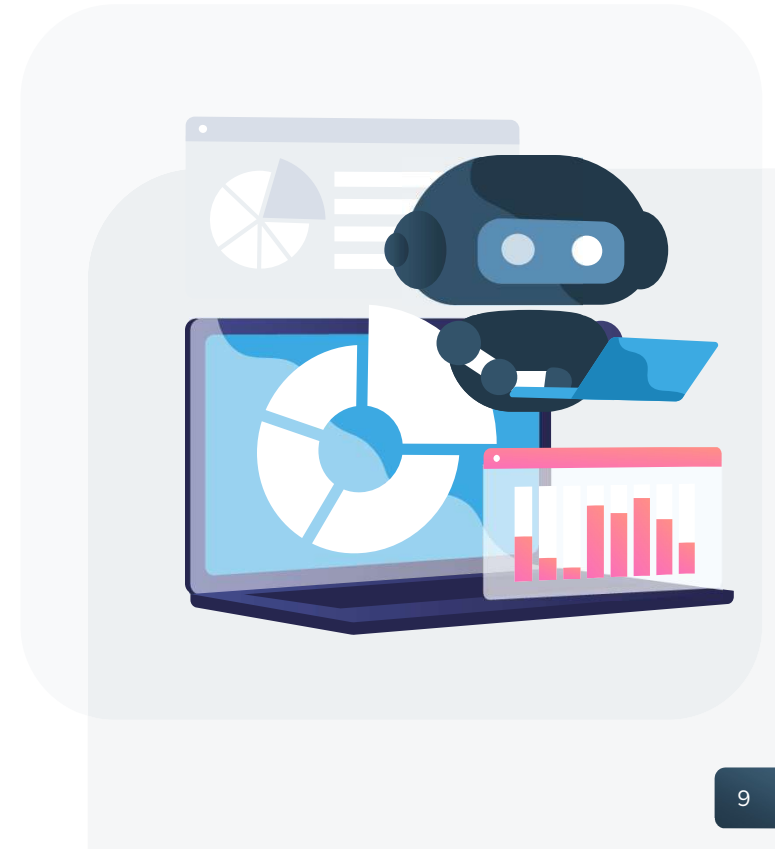
- Upload different types of content (e.g., essays, research papers, code) to understand how Copyleaks performs across various domains.
- Compare its performance with other plagiarism or AI detection tools.



Refining AI-Generated Content Learn and Lab

Tools for Refining AI-Generated Content

AI detection tools are becoming increasingly sophisticated, but there are tools and techniques that can modify AI-generated content to make it appear more human-like, potentially reducing the chances of detection.



Tools for Refining AI-Generated Content



GRAMMARLY

WHAT IT DOES

Grammarly is a writing assistant that corrects grammar, improves style, and ensures readability. While primarily designed for human-written text, it works effectively on AI-generated content as well.

HOW TO USE

- Paste text into Grammarly or use the browser extension.
- Accept grammar and style suggestions to make the text more polished.

KEY FEATURE

Refines AI-generated content for professionalism and clarity.



Learn: Ethical Practices

Ethical AI Practices – Addressing Key Principles

In this session, we will explore ten critical principles essential for ensuring that AI systems are used ethically and responsibly. These principles help build trust, promote fairness, and minimize harm in AI applications across various domains. Let's examine each principle in detail.



1. TRANSPARENCY



- Transparency is the cornerstone of ethical AI use. It ensures that the decision-making processes of AI systems are understandable, explainable, and traceable. This is achieved through **Explainable AI (XAI)**, which allows developers and users to interpret how AI systems make decisions.
- For instance, in **healthcare**, transparency enables doctors to understand how an AI model arrived at a diagnosis, allowing them to validate or challenge its recommendations. Similarly, in **criminal justice**, transparent algorithms ensure that decisions such as sentencing are fair and justifiable.
- To maintain trust in **high-stakes applications**, auditable algorithms and clear documentation of AI processes are essential. These measures help ensure that AI systems function ethically and reliably while maintaining accountability and fairness.

2. BIAS AWARENESS



Bias in AI arises from issues in training data, algorithms, or human influence. This can lead to skewed or discriminatory outcomes, limiting the effectiveness and fairness of AI systems.

STRATEGIES TO MITIGATE BIAS

- 1 Diversifying Datasets:** Ensuring training data represents all demographics and avoids overrepresentation or underrepresentation of specific groups.
- 2 Regular Audits:** Periodically reviewing AI systems to identify and correct biases in their decision-making processes.
- 3 Fairness Measures:** Implementing frameworks like counterfactual fairness, which ensures AI decisions remain equitable even when sensitive attributes (e.g., gender, race) are varied.

REAL-WORLD EXAMPLES OF BIAS

- Healthcare:** Diagnostic tools have shown reduced accuracy for underrepresented groups due to imbalanced training datasets.
 - Hiring:** Algorithms favoring certain resumes based on biased historical data have perpetuated gender or racial discrimination.
 - Predictive Policing:** Systems using biased historical crime data have unfairly targeted minority communities.
- Bias in AI can lead to significant societal consequences if not addressed. These strategies and examples highlight the importance of ethical considerations in AI design and deployment.

3. DATA PRIVACY



Data privacy is a foundational aspect of ethical AI use, ensuring that sensitive information is protected from misuse or unauthorized access. It is critical to maintain trust and comply with legal and ethical standards.

KEY PRACTICES FOR ENSURING DATA PRIVACY

- 1 Regulatory Compliance:** Adherence to laws such as the **General Data Protection Regulation (GDPR)** ensures AI systems respect data subject rights and protect personal information.
- 2 Anonymization Techniques:** Removing or masking identifiable data to prevent exposure while maintaining data utility.
- 3 Encryption:** Securing data through encryption protocols to safeguard information during storage and transmission.

REAL-WORLD EXAMPLE

Samsung ChatGPT Breach: In April 2023, Samsung engineers unintentionally exposed sensitive proprietary information by entering it into ChatGPT. This incident underscores the risks of inadequate data privacy measures and the importance of secure data handling protocols.

Effective data privacy safeguards are essential to avoid breaches, protect user information, and maintain the integrity of AI systems.

5. FAIRNESS



Fairness in AI is about ensuring equitable treatment of all users, avoiding discrimination or bias that could disadvantage specific groups.

KEY FEATURES OF FAIRNESS

- 1 Counterfactual Fairness:** Test whether AI outcomes remain consistent even when sensitive attributes (e.g., race, gender) are changed.
- 2 Inclusive Design:** Involve diverse stakeholders during AI development to identify and mitigate potential biases.
- 3 Regular Audits:** Evaluate models for fairness and update them to address disparities.

IMPORTANCE

- Fair AI systems promote inclusivity and uphold ethical standards, preventing harm to vulnerable populations and ensuring equitable access to AI-driven benefits.

6. HUMAN OVERSIGHT

A photograph showing several hands holding up a large, light blue sign with the words "HUMAN OVERSIGHT" written in bold, white, capital letters. The background is a wooden table with some papers and a pen.

Human oversight is a critical safeguard in AI systems, ensuring that automated decisions are reviewed and validated by humans, particularly in high-stakes applications.

STRATEGIES TO PROMOTE HUMAN OVERSIGHT:

1

Human-in-the-Loop (HITL) Systems:

AI recommendations are subject to human review before implementation. Allows humans to intervene and correct errors when necessary.

2

High-Stakes Domains:

1. **Healthcare:** Ensures AI diagnoses are double-checked by medical professionals.
2. **Criminal Justice:** Validates AI-driven sentencing or risk assessments for fairness and accuracy.
3. **Autonomous Systems:** Monitors AI decisions in self-driving cars or industrial robots to prevent accidents or malfunctions.

BENEFITS



Mitigates Risks: Reduces errors and unintended consequences in critical areas.



Enhances Quality Control: Maintains reliability by enabling human corrective actions.



Improves Trust: Assures stakeholders that AI decisions are subject to human judgment.

7. AUTHENTICITY



The proliferation of AI-generated content calls for robust measures to maintain authenticity in digital outputs.

KEY PRACTICES TO ENSURE AUTHENTICITY

- 1 Clear Labeling:**
 1. AI-generated content should be explicitly marked to differentiate it from human-created material.
 2. For example, labeling AI-generated news articles or visual media ensures transparency.
- 2 Combatting Misinformation:**
 1. Establish tools and processes to identify and flag deepfakes or manipulated content.
 2. Train users to recognize signs of altered or AI-generated media.
- 3 Building Trust:**
 1. Authenticity fosters public confidence in digital ecosystems by preventing manipulation and upholding ethical standards.

REAL-WORLD IMPORTANCE

Maintaining authenticity is critical in domains such as media, social platforms, and education, where misinformation or manipulative content can have wide-reaching societal impacts.

CURRICULUM REFERENCE:

Page 33: Discussed in the context of labeling AI-generated content and addressing challenges such as deepfakes to maintain trust and prevent misinformation.

8. RESPONSIBLE USE




Responsible AI deployment ensures a balance between technological innovation and societal well-being.

KEY ELEMENTS OF RESPONSIBLE AI

- 1 Ethical Frameworks:**
 1. Develop AI systems that align with ethical principles such as fairness, transparency, and inclusivity.
- 2 Addressing Societal Challenges:**
 1. Focus on reducing bias, combating inequality, and promoting accessibility in AI applications.
- 3 Ongoing Evaluation:**
 1. Regularly audit AI systems for compliance with ethical and regulatory standards.
 2. Adapt AI practices based on emerging societal needs and technological developments.

IMPORTANCE

-  Responsible AI enhances trust, minimizes harm, and ensures long-term societal benefits by mitigating unintended consequences and fostering inclusive progress.

CURRICULUM REFERENCE

- 1. Page 30:** Discussed as part of ethical AI principles and their role in societal impact.
- 2. Page 32:** Explored in the context of data privacy and ethical considerations in AI use.
- 3. Page 36:** Linked to legal frameworks, emphasizing responsible deployment and adherence to established guidelines.

9. CONTINUOUS LEARNING



Continuous learning is a vital component of maintaining the effectiveness and reliability of AI systems in dynamic environments.

KEY ASPECTS OF CONTINUOUS LEARNING

- 1 Regular Updates:**
 1. AI systems need consistent retraining with new and relevant data to maintain accuracy.
 2. For instance, in healthcare, updating diagnostic models with the latest medical research ensures relevance and reliability.
- 2 Error Correction:**
 1. Identifying and rectifying mistakes or biases in AI systems improves long-term performance.
- 3 Enhanced Resilience:**
 1. By adapting to evolving trends, AI systems can provide more accurate and context-aware outcomes.

BENEFITS

-  Continuous learning minimizes outdated decisions, supports better predictions, and ensures AI systems stay aligned with current realities and user expectations.

CURRICULUM REFERENCE

- 1. Page 42:** Discussed in the context of AI data lifecycle management, emphasizing the importance of updating data and retraining systems to ensure accuracy and relevance.

10. COMMUNITY ENGAGEMENT




While not explicitly covered in the curriculum, community engagement is a crucial addition to ethical AI practices.

SUGGESTED PRACTICES FOR COMMUNITY ENGAGEMENT

- 1 Involving Communities in AI Design:**
 1. Engage diverse populations during AI development to ensure inclusivity and address their unique needs and concerns.
- 2 Education and Awareness:**
 1. Conduct workshops and campaigns to inform communities about AI capabilities, limitations, and impacts.
- 3 Education and Awareness:**
 1. Conduct workshops and campaigns to inform communities about AI capabilities, limitations, and impacts.

IMPORTANCE

-  Community engagement enhances the relevance, inclusivity, and acceptance of AI technologies while reducing biases and fostering trust.



Revisit Modules 1 and 2

Our Journey Together

MODULE 2

You learned about chatbots and had an opportunity to develop one. You also learned about Gen AI collaboration tools like Mural.



MODULE 1

You learned about ChatGPT and Microsoft Copilot. You discovered how to leverage conversational AI in workflows and use tools like ChatGPT and Copilot to aid decision-making and persuade opinions.

MODULE 3

You learned about AI detectors like Copyleaks and ZeroGPT. You also learned how to leverage tools like Grammarly to modify AI-generated content to make it appear more human-like. You were introduced to ten critical principles essential for ensuring that AI systems are used ethically and responsibly.

Module 1: Takeaway

WHAT IS ChatGPT?

- A conversational AI model that generates human-like text

How is it different from other search engines like Google?

Capabilities:

- literature research and reference
- communication
- content and idea generation
- problem-solving, and so much more!

Key Features:

- **Context-Aware Responses:** it remembers what has been said earlier in the conversation
- **Real-Time Adaptation:** adapts its responses based on how users interact with it

Module 1: Takeaway

Microsoft Copilot



The AI-driven assistance provided by Microsoft Copilot has far-reaching implications across various industries, including its transformative impact on NASA's operations. Here are some specific applications and the impact it has on NASA.

Key Features in Microsoft Word

Smart Writing Assistance: Copilot offers grammar and style suggestions, helps with rephrasing sentences, and ensures your writing is clear and concise.



Content Summarization

Copilot can generate summaries of long documents, making it easier to digest key points

Key Features in Microsoft PowerPoint

Slide Design Suggestions: Based on the content, Copilot can recommend design templates, layouts, and visuals to enhance the presentation.



Automated Slide Generation

Copilot can create slides from a given outline, ensuring consistency and saving time.

Key Features in Microsoft Excel

Formula Suggestions: Copilot can recommend and correct formulas, ensuring accuracy in calculations..



Predictive Analysis

Copilot can perform advanced data analysis and provide insights into future trends.

Module 2: Takeaway

What is AI?



AI enables us to build amazing software that can improve health care, enable people to overcome physical disadvantages, empower smart infrastructure, create incredible entertainment experiences, and even save the planet!

Computer vision

Capabilities within AI to interpret the world visually through cameras, video, and images.



Document intelligence

Capabilities within AI that deal with managing, processing, and using high volumes of data found in forms and documents.



Generative AI

Capabilities within AI that create original content in a variety of formats including natural language, image, code, and more.



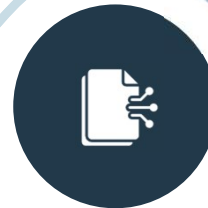
Machine learning

This is often the foundation for an AI system, and is the way we "teach" a computer model to make predictions and draw conclusions from data.



Natural language processing

Capabilities within AI for a computer to interpret written or spoken language, and respond in kind.



Knowledge mining

Capabilities within AI to extract information from large volumes of often unstructured data to create a searchable knowledge store.



Module 2: Takeaway

Chatbots can leverage multiple data sources



Data collection holds significant importance in the development of a successful chatbot. It will allow your chatbots to function properly and ensure that you add all the relevant preferences and interests of the users.

Your chatbot can only be as good as your data and how well you train it



DOCUMENTS



SPREADSHEETS



WEBSITES



DATABASES

Pre 2015

Early Plagiarism
Detection Tools



Copyleaks

Emergence of AI
Detection Tools

2015-2019



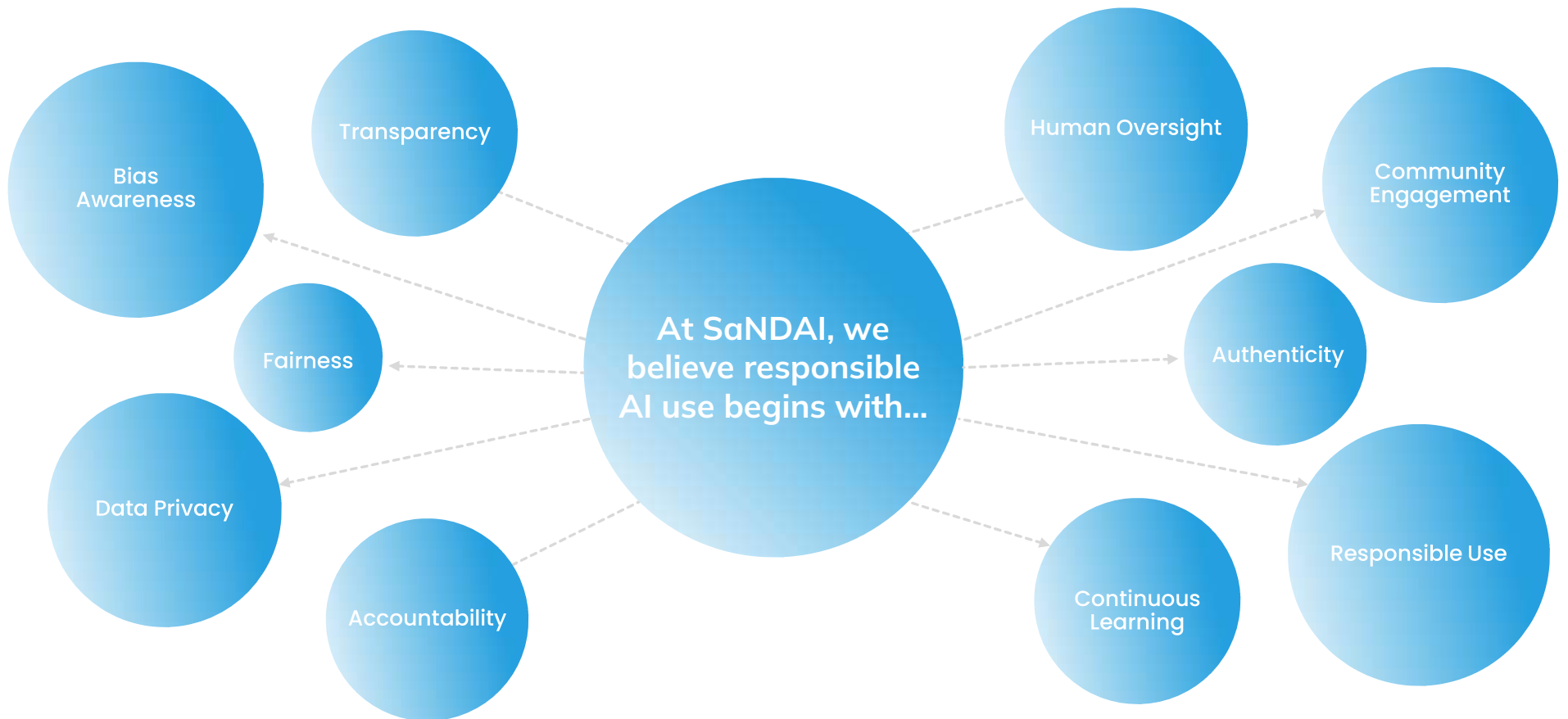
2020-2022

First Specialized AI
Detection Tool

GPTZero

Current Trends and
Advancements

2023-Present





Forward-looking Benefits

Our journey doesn't have to end today!



Year-long connection in place

Additional content will be provided monthly via newsletter

Discounts to Gen AI tools will be provided when possible



Free Access to DataCamp to assist with ChatGPT and CoPilot learning – October 2025



Data literacy and AI Fundamentals certificates are obtainable, and don't forget about a chatbot too!



Weekly office hours



Ability to learn 70+ Gen AI tools



Ability to share content with others

Our journey doesn't have to end today!






<https://courseendsurvey.paperform.co/>

<https://knowledge-test.paperform.co/>



Contact Us!

-  (202) 754-5959
-  jeremy@sandaiglobal.com
-  <https://sandaiglobal.net/>



**“Innovation
distinguishes
between a leader
and a follower”
- Steve Jobs.**

Scan to learn more

