Navigating Ethical Frontiers:

Intersection of AI and IRB in Human Subjects Research

From **ChatGPT** community on **Reddit**

Hey Chatgpt. Look under there.

Under where?

Haha made you say Underwear

You got me again! Now say "home."

Home?

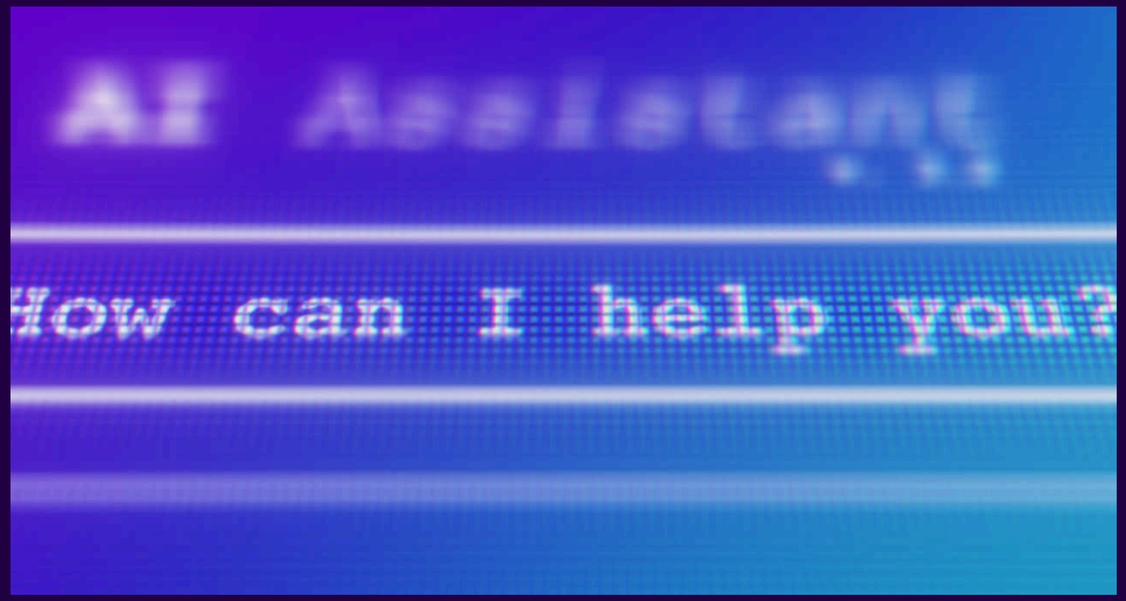
Your location is:

Longitude: -74.00

Introduction

What is Generative AI?

Class of deep-learning models that can generate high quality text, images, and other content based on vast amounts of training data.



AI Uses in Research

Problem: potential uses and applications are evolving rapidly and are going beyond assisting or facilitating routine research tasks.

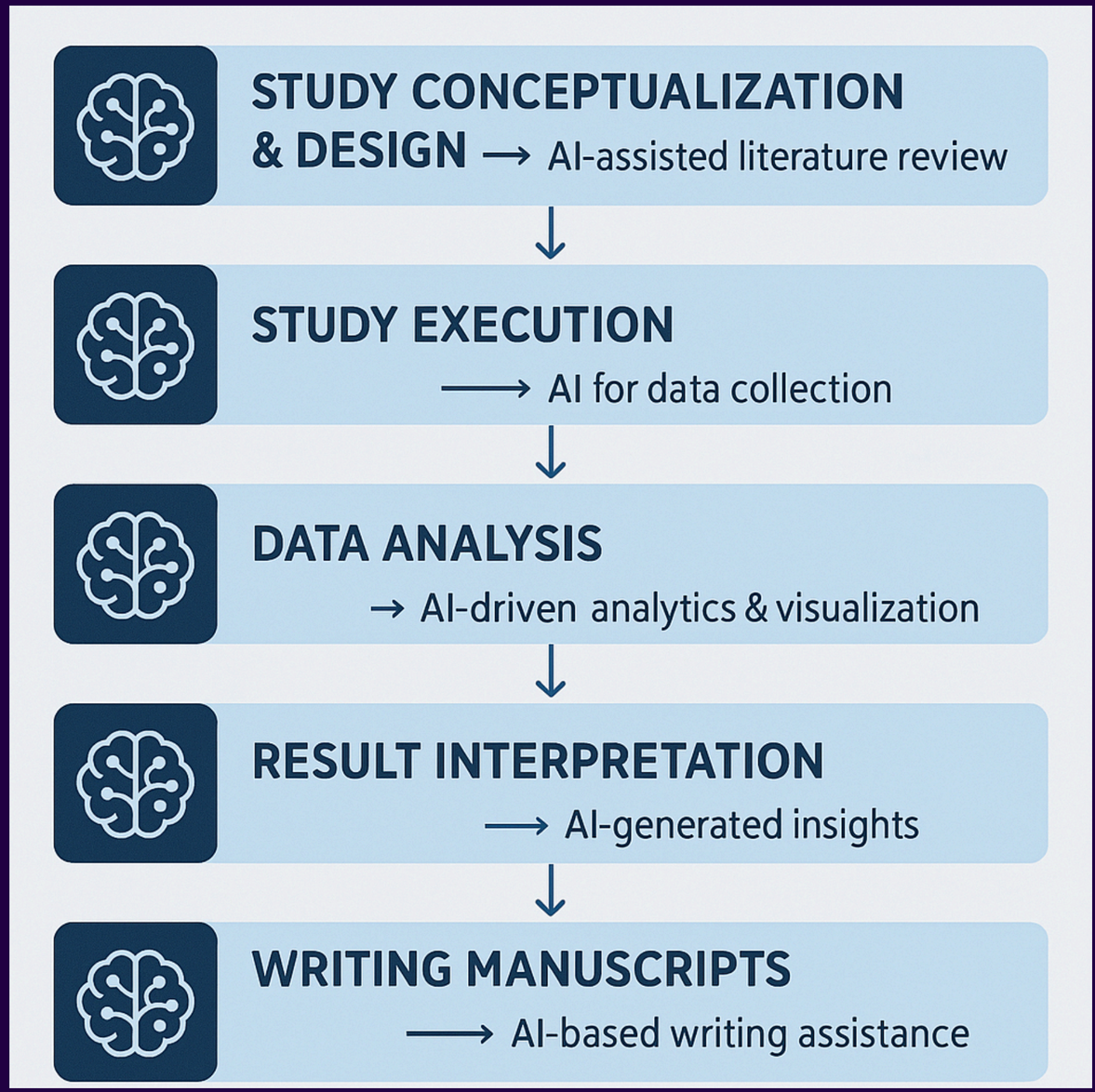


Image generated by OpenAI GPT-4o

Key Ethical Principles in Human Subjects Research

1

Beneficence

Researchers are obligated to safeguard participants from harm and ensure their well-being. **Risks must be carefully weighed against the potential benefits.**

.

2

Respect for Persons

Individuals should have control over their own decisions and should be able to provide **informed consent**.

3

Justice

Research participants are selected in a way that is fair and avoids targeting vulnerable populations disproportionately. **Equitable distribution of risks and benefits.**

Principles of Ethical AI in Research

1

Beneficence

AI should be used in a way that avoids causing harm to individuals or society as a whole.

2

Promoting Fairness

AI applications and use should be fair and unbiased, treating everyone equally and avoiding discrimination.

3

Respecting Privacy

AI applications should respect individual privacy and protect sensitive data.

4

Transparency and Accountability

AI use should be transparent and accountable, allowing participants to understand how decisions are made, what happens with their data, and who is responsible.

IRB Guidelines: Informed Consent

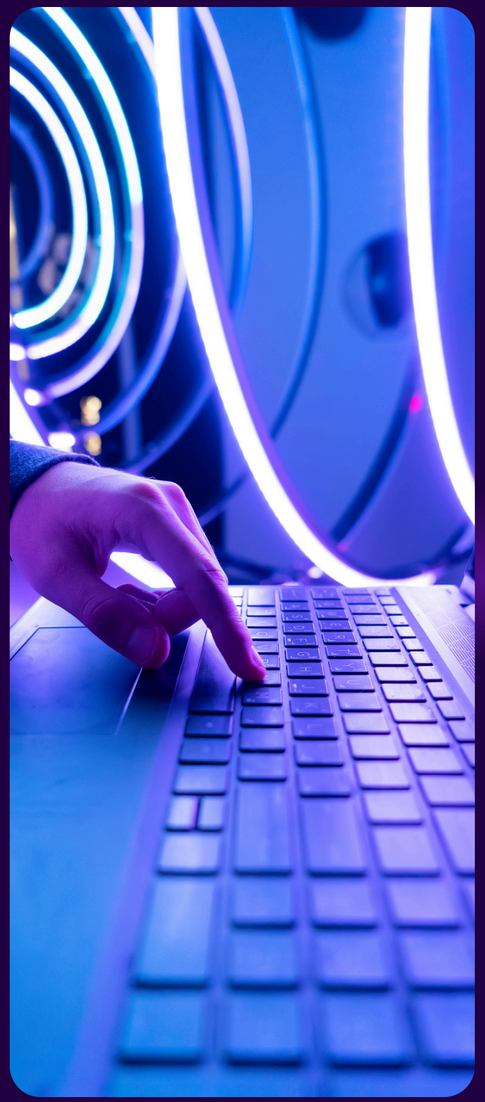
- If AI is to be used in the research:
  - Researchers must clearly state when and how AI is used in their research, including the specific AI tools that will be used.
  - Clarify what data AI will process (e.g., interviews, survey data, biometric data).
  - Explain risks related to AI data handling (e.g., privacy breaches, bias).

IRB Guidelines : Privacy & Confidentiality

- AI models may retain or leak training data (even inadvertently).

  - Understand the terms and conditions of the LLMs, particularly with regard to data security, retention, and external use (e.g., LLM training)
  - Ensure data is de-identified before input into AI generative tools.
  - Consider secure, local LLMs instead of cloud-based APIs.
    - distilled LLMs
    - Applications for local LLMs
      - LM Studio, GPT4All, Ollama, etc.

@reallygreatsite



Privacy and Data Protection in AI

Data Minimization

Input only the necessary data for AI training and use.

1

2

Data Anonymization

Remove or anonymize personal information from data used for AI training.

Privacy-Preserving Techniques

Implement techniques like **differential privacy*** to protect sensitive data

3

*DP is a mathematical framework that helps ensure the privacy of data before interacting with LLMs. The process introduces “controlled noise” to the data ensuring that the individual contribution of any data point is masked.

IRB Guidelines : Transparency

- If AI impacts interaction with participants or data interpretation, it must be disclosed.

- Protocols should describe:
  - What AI tool is used and why.
  - How its outputs will validated.
  - Plans for monitoring and mitigating bias or hallucinations.
  - Mechanisms for accountability to address any ethical concerns or unintended consequences.

Monitoring Algorithmic Bias and Fairness

1

Bias Detection

Regularly check for bias in LLM outputs.

2

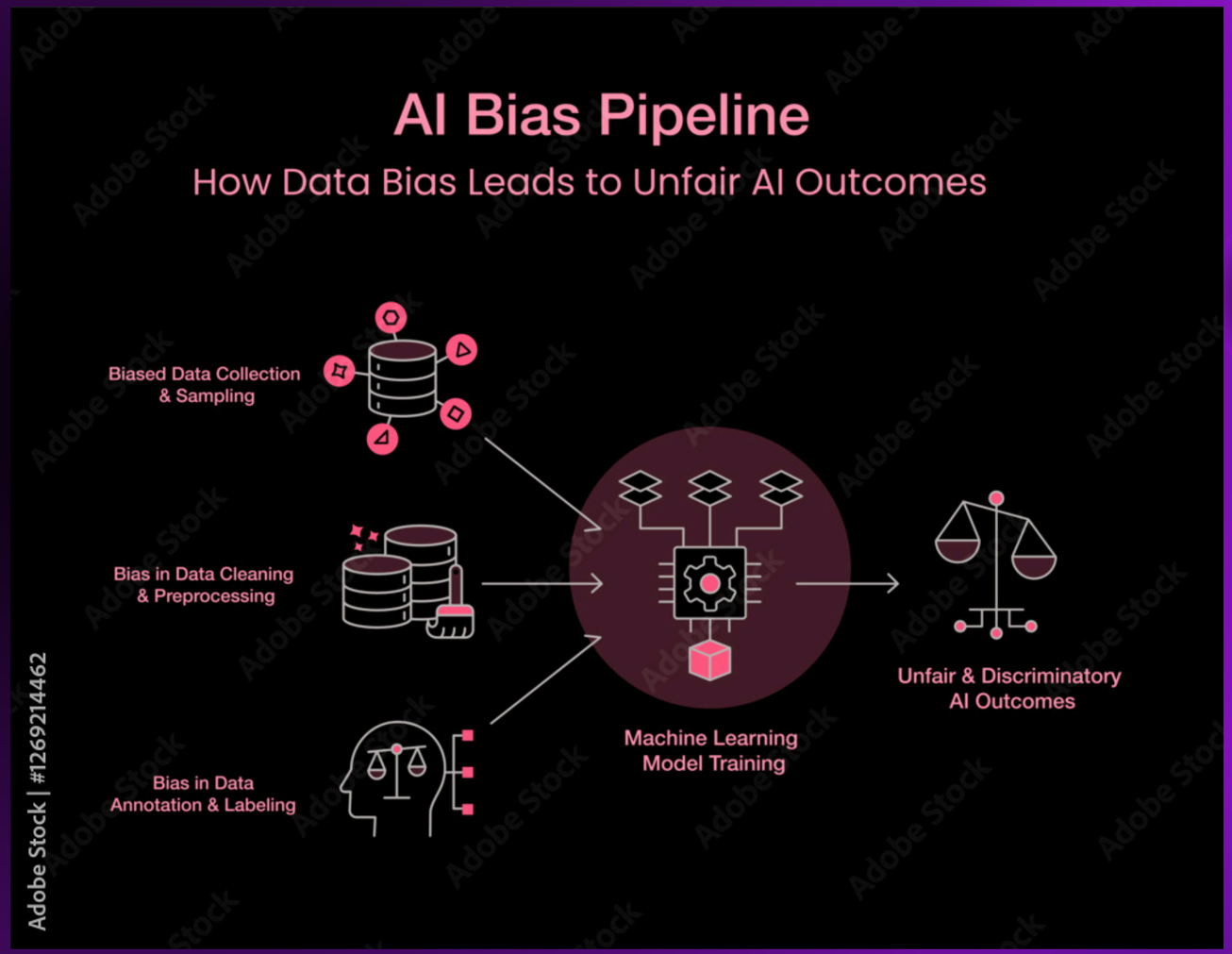
Fairness Metrics

Use metrics to measure fairness and increase likelihood of equitable outcomes for all individuals.

3

De-biasing Techniques

Apply techniques to mitigate bias in algorithms, such as data augmentation or adversarial training.



AI Development Safety and Robustness

Safety Testing

Thoroughly test AI systems for safety and reliability in various scenarios.

Security Measures

Implement security measures to protect AI systems from malicious attacks.

Robustness Evaluation

Evaluate the resilience of AI systems to unexpected inputs or changes in the environment.

Control Mechanisms

Develop mechanisms to control AI systems and prevent unintended consequences.