

# Key Regulatory Terms and Artificial Intelligence (AI) Tools in Human Subjects Research

Iris L. Jenkins, PhD

Director, Research Integrity and Consultation

OHRP Exploratory Workshop

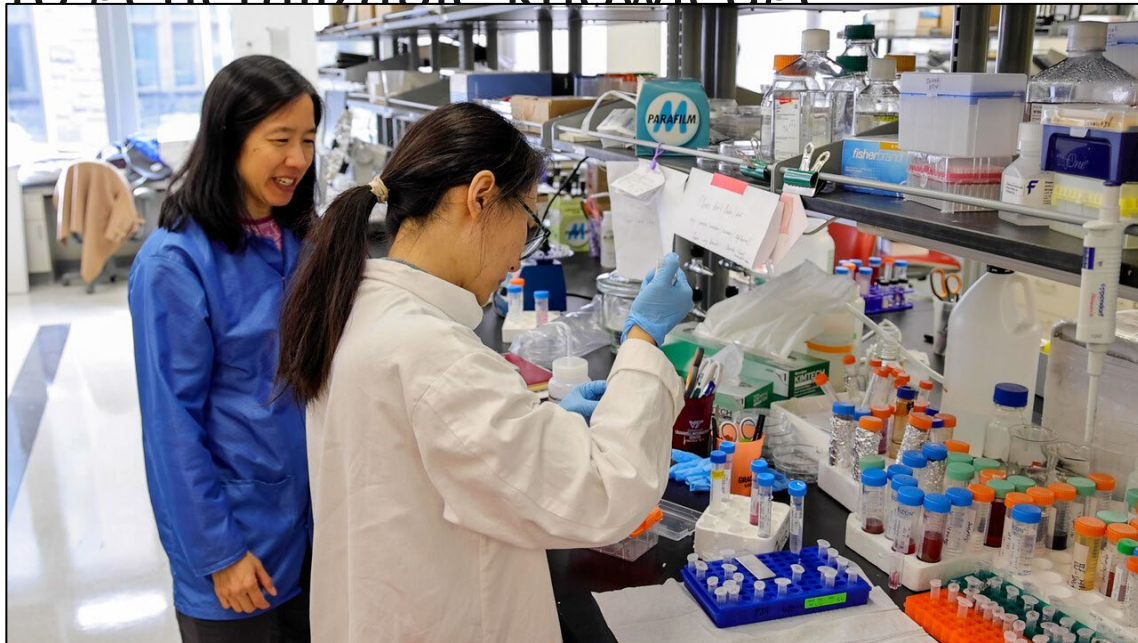
September 2024



RESEARCH AND INNOVATION  
SCHOLARLY INTEGRITY AND  
RESEARCH COMPLIANCE  
VIRGINIA TECH.

# Regulatory Definitions (45 CFR 46)

**Research** means a systematic investigation, including research development, testing, and evaluation, designed to develop or contribute to generalizable knowledge

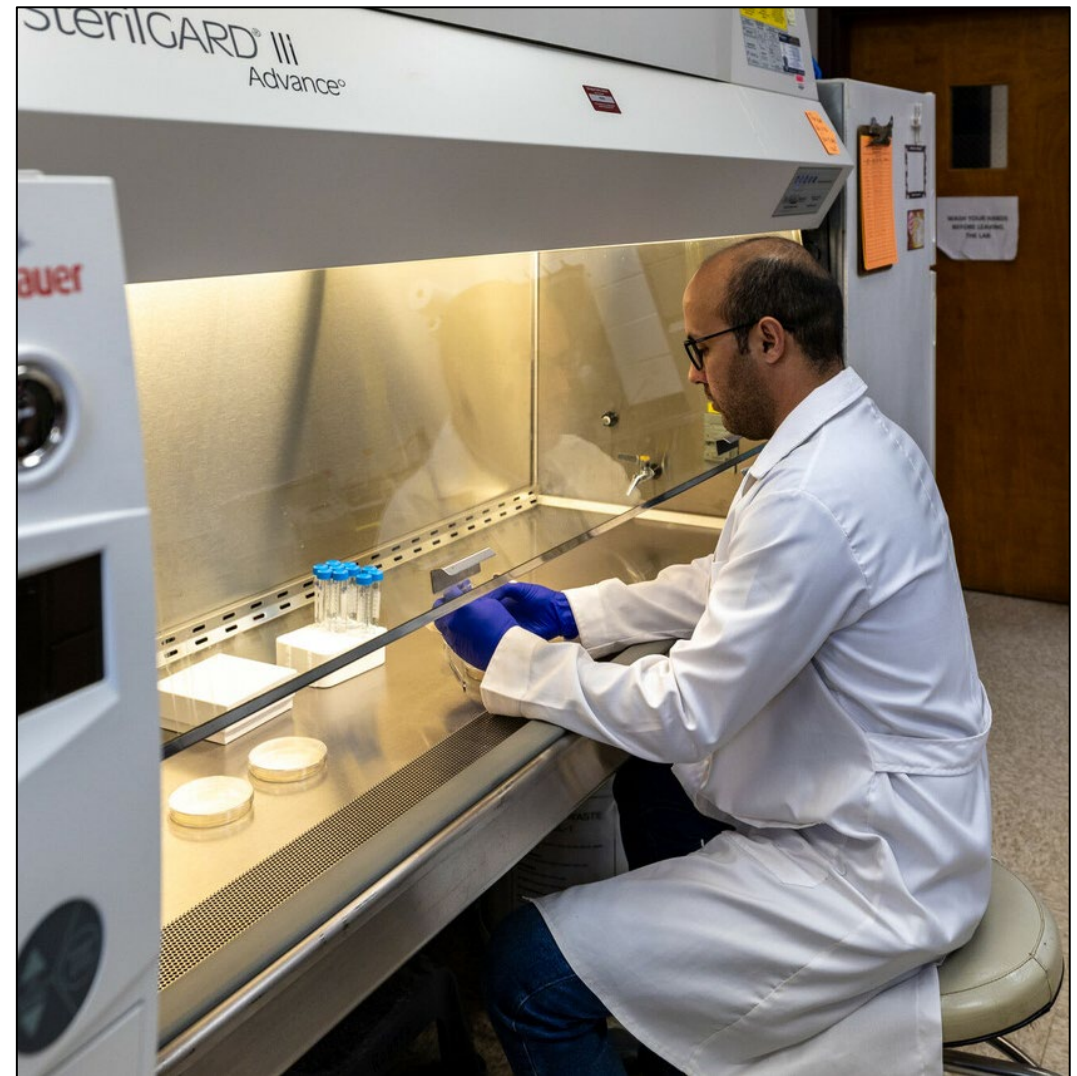




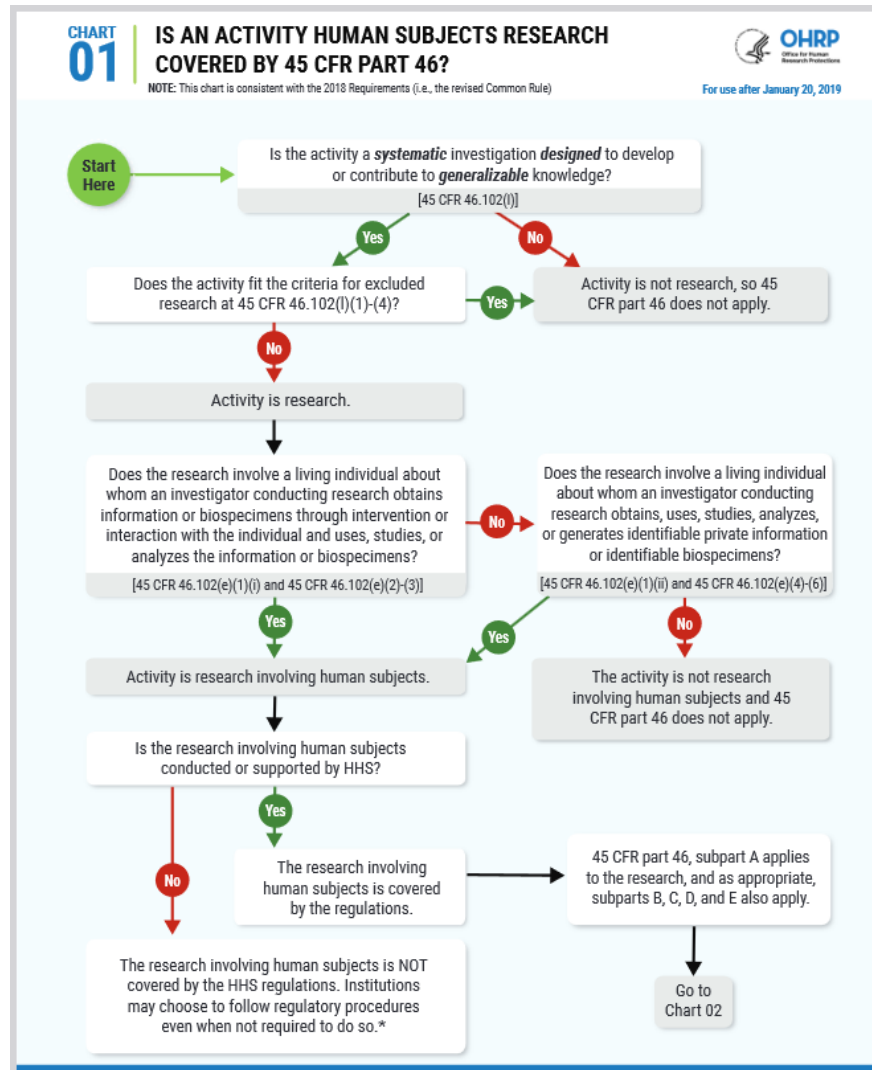
## Regulatory Definitions (45 CFR 46)

***Human subject*** means a living individual about whom an investigator (whether professional or student) conducting research:

- (i) Obtains information or biospecimens through intervention or interaction with the Individual, and uses, studies, or analyzes the information or biospecimens; or
- (ii) Obtains, uses, studies, analyzes, or generates identifiable private information or identifiable biospecimens.



# Human Subjects Research Determinations



## ■ Is the activity research?

- Is the activity systematic?
- Is the activity designed to develop or contribute to generalizable knowledge?

## ■ Does the activity involve human subjects?

- About whom?
- Information or biospecimens obtained through intervention or interaction?
- Identifiable private information?
- Identifiable biospecimens?

# Scenario #1

Proposal to evaluate the effects of television viewing on food intake.

During one session participants will be provided a meal to eat while viewing television and during another session they will be provide a meal to eat without television. Participants will also complete demographic questionnaires.

Measure the amount of food eaten by participants and compare the amounts across sessions.

## **Research**

- Systematic
- Designed to contribute to generalizable knowledge

## **Involves human subjects**

- About the behavior of people
- Information obtained by intervention/interaction
- May include identifiable private information

# Scenario #2

Proposal to evaluate the impact of an **intervention by a robot** on food intake during television viewing.

Participants will have an educational **intervention provided by a robot** prior to being provided a meal to eat while viewing television. Participants will also complete demographic questionnaires.

Measure the amount of food eaten by participants and compare the amounts across those with robot intervention and those without.

## Research

- Systematic
- Designed to contribute to generalizable knowledge

## Involves human subjects

- About the behavior of people
- Information obtained by intervention/interaction
- May include identifiable private information

## Scenario #3

Proposal to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information.

Involves feeding previously collected, de-identified data into an AI tool.

A bit more complicated...

# Scenario #3

Proposal to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information.

Feed previously collected, de-identified data into an AI tool.

## A bit more complicated...

### Is it research?

- **Systematic** - yes
- **Designed to contribute to generalizable knowledge** - investigators state there is no intention to generalize at this point – only train an algorithm



# Scenario #3

Proposal to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information.

Feed previously collected, de-identified data into an AI tool.

## A bit more complicated...

### Does it involve human subjects?

- **About whom?** - Investigators argue it is not about any person, but development of an algorithm
- **Information obtained by intervention/interaction** – secondary data analysis, no intervention/interaction
- **Identifiable private information?** - data are de-identified

# Scenario #4

Proposal to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information.

Provided with identifiable information, which will be de-identified and fed into an AI tool.

## A bit more complicated...

### Is it research?

- **Systematic** - yes
- **Designed to contribute to generalizable knowledge** - investigators state there is no intention to generalize at this point – only train an algorithm

# Scenario #4

Proposal to train an algorithm to predict the amount of food a person will eat while viewing television based on various demographic information.

Provided with identifiable information, which will be de-identified and fed into an AI tool.

## A bit more complicated...

### Does it involve human subjects?

- **About whom?** - Investigators argue it is not about any person, but development of an algorithm
- **Information obtained by intervention/interaction** – secondary data analysis, no intervention/interaction
- **Identifiable private information?** - initially information is identifiable

# If it is human subjects research...Exempt #4?

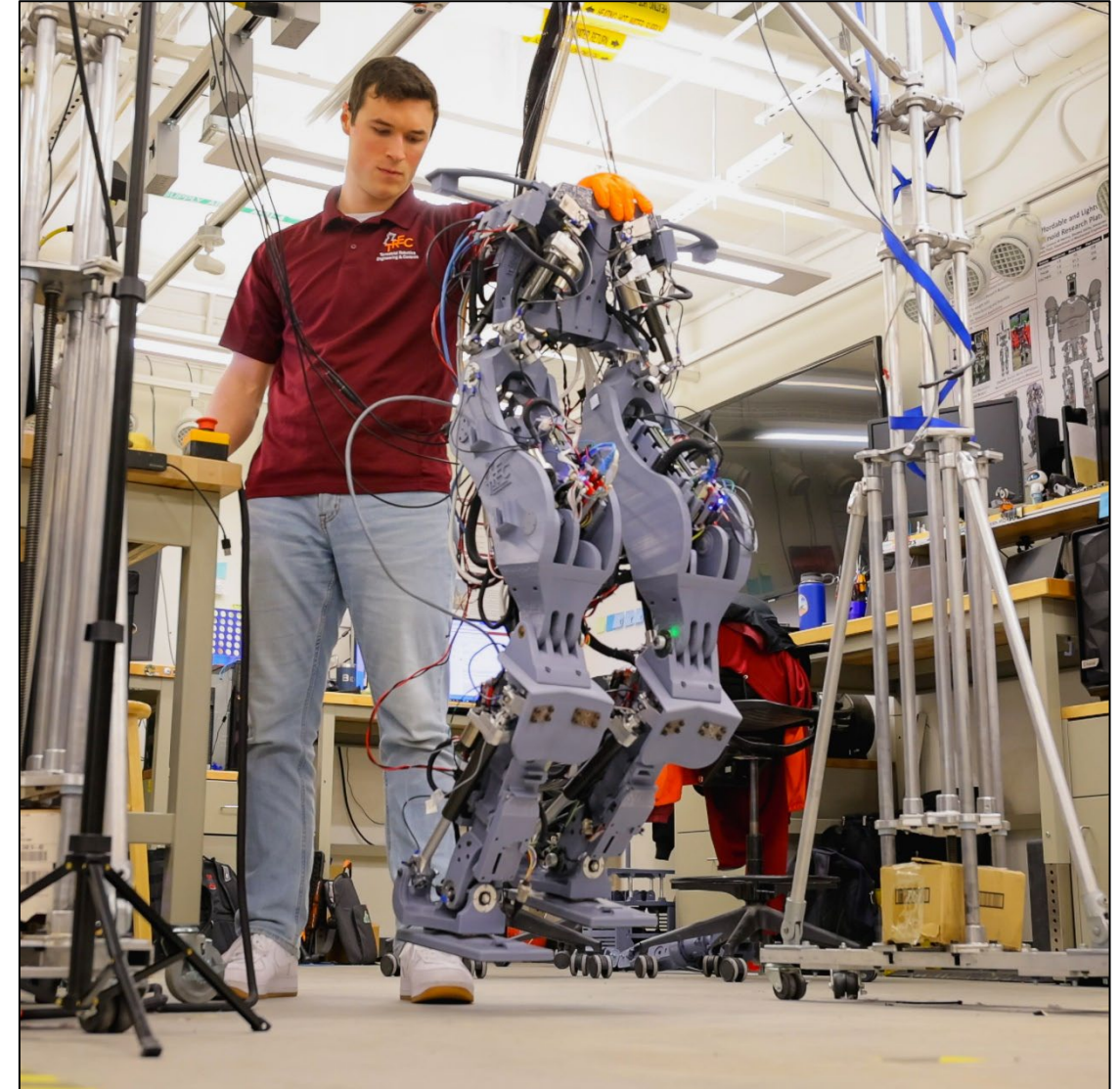
Secondary research for which consent is not required: Secondary research uses of identifiable private information or identifiable biospecimens, if at least one of the following criteria is met:

- 1.(i) The identifiable private information or identifiable biospecimens are publicly available;
- 2.**(ii) Information, which may include information about biospecimens, is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained directly or through identifiers linked to the subjects, the investigator does not contact the subjects, and the investigator will not re-identify subjects;**
- 3.(iii) The research involves only information collection and analysis involving the investigator's use of identifiable health information when that use is regulated under 45 CFR parts 160 and 164, subparts A and E, for the purposes of "health care operations" or "research" as those terms are defined at 45 CFR 164.501 or for "public health activities and purposes" as described under 45 CFR 164.512(b); or
- 4.(iv) The research is conducted by, or on behalf of, a Federal department or agency using government-generated or government-collected information obtained for nonresearch activities, if the research generates identifiable private information that is or will be maintained on information technology that is subject to and in compliance with section 208(b) of the E-Government Act of 2002, 44 U.S.C. 3501 note, if all of the identifiable private information collected, used, or generated as part of the activity will be maintained in systems of records subject to the Privacy Act of 1974, 5 U.S.C. 552a, and, if applicable, the information used in the research was collected subject to the Paperwork Reduction Act of 1995, 44 U.S.C. 3501 et seq.



# Summary

- Many AI projects will not readily fit the definition of “human subjects research”
- Many AI projects meeting the definition of “human subjects research” fit into the “Exempt” category → limited oversight
- Concerns remain:
  - Privacy and confidentiality when data are fed into AI
  - Individual de-identified datasets combined with other data sets → identifiability
  - Determinations for publicly available data (what is “publicly available”?)





RESEARCH AND INNOVATION  
**SCHOLARLY INTEGRITY AND  
RESEARCH COMPLIANCE**  
VIRGINIA TECH.