Northwestern
INFORMATION TECHNOLOGY

Incorporating AI Into Your Research Responsibly

IRB Brown Bag

Christina Maimone
Northwestern IT Research Computing and Data Services

Daniel Schneider
NMEDW

April 2026

See also from May 2024:

Responsible Use of Artificial Intelligence in Human Research

https://bit.ly/IRB-AI-May2024

Research Computing and Data Services

https://bit.ly/rcdsinfo

- Platforms and services in support of all researchers
- Computing (Quest, Genomics Compute Cluster)
- Data management (storage, security, workflows)
- Data science, AI, statistics, and visualization

Northwestern Medicine Enterprise Data Warehouse(NMEDW)

- The NMEDW is a joint initiative across the Northwestern University Feinberg School of Medicine (FSM) and Northwestern Memorial HealthCare (NMHC)
- It is a single, comprehensive, integrated repository of all clinical and research data sources on the campus to facilitate research, clinical quality, healthcare operations, and medical education

Note: AI services and guidance are changing over time. This information is from April 2026.

it.northwestern.edu/ai

Security:
What is IT worried about?

Confidentiality of Research

Think of sharing data with a service/platform/ app/company like sharing with a person

- Should that entity have access to your research data?
- Do they know how, and have the right tools, to protect it?
- What is the recourse if something goes wrong?

Data Classification Levels

Level 1: Public data: Is public or OK to be public

Level 2: Sensitive, not public: General research data, university business operations

Level 3: Contractual and legal restrictions: Personal sensitive information (SSNs), PHI, DUAs, restricted financial data, CUI, NIST SP 800-171

~~**Level 4: Classified, Export Controls**~~

FSM: https://www.feinberg.northwestern.edu/it/docs/fsm_data_storage_policy_v2.pdf
General: https://policies.northwestern.edu/docs/data-classification-policy.pdf

Data Exposure

Data being stolen or publicly exposed through attacks on, mistakes by, or insufficient security practices of third-party companies or tools

Private **data leaking** into the public by being used to train new AI models

Reidentification of data when combined with other information

Contracts

Rapidly **changing user agreements** and terms for AI services

Ownership/transfer/retention of data: inputs, outputs, prompts, results

Maintaining terms of **data use agreements**

AI Agents

Broad **access to file systems** may exacerbate problems with overly liberal permissions settings

Unsupervised agents misusing University systems

Credentials leaking through hacks to open source tools

Widescale **systematic security attacks** exploiting vulnerabilities in research applications

Use University-Approved Resources

Do not upload research data into any tool, service, or website (AI-related or not) that is not approved by the university for use with human research data

Northwestern AI Tools

Microsoft Copilot

Copilot Chat (Free)
M365 Copilot (Premium) ~$18/month/user

- Has University data protections; appropriate for level 2 and some level 3 data
- Best choice for chat-based tool for human research data
- Notably improved in the past few months

Cloud Platforms

Microsoft Azure, Amazon Web Services (AWS), Google Cloud (GCP)

Primary: API access to LLMs and other generative AI models
Secondary: Other AI-enabled services

All have AI models where the data stays within Northwestern's environment and stays protected by University contracts; can require additional security controls

AI Transcription

Video and audio recordings of research participants

1. Northwestern Zoom/Teams (no HIPAA data)
2. Software that runs locally on your computer without an internet connection
3. Services in Northwestern Cloud platforms

Writing Code

- Be careful when you have sensitive data!
- Working in a chat interface is safest
- Options for assistants/agents with sensitive data:
  - GitHub Copilot Enterprise (per user cost)
  - Using APIs set up through Northwestern cloud providers (only possible with some tools, usage-based cost)
  - Local models: useful models generally require hardware beyond a laptop

Choosing AI Tools for Research Protocols

Does it run locally on your computer, or does it send data over the internet?

What information will it have access to?

Who To Talk To

Medical Records Data: NMEDW Research Analytics https://www.feinberg.northwestern.edu/it/services/research-analytics-edw/

Other Feinberg Research: FSM IT fsmhelp@northwestern.edu

Everyone Else: Northwestern IT Research Computing and Data Services researchdata@northwestern.edu

Research Use Guidelines

Use AI tools to help you, not to think for you

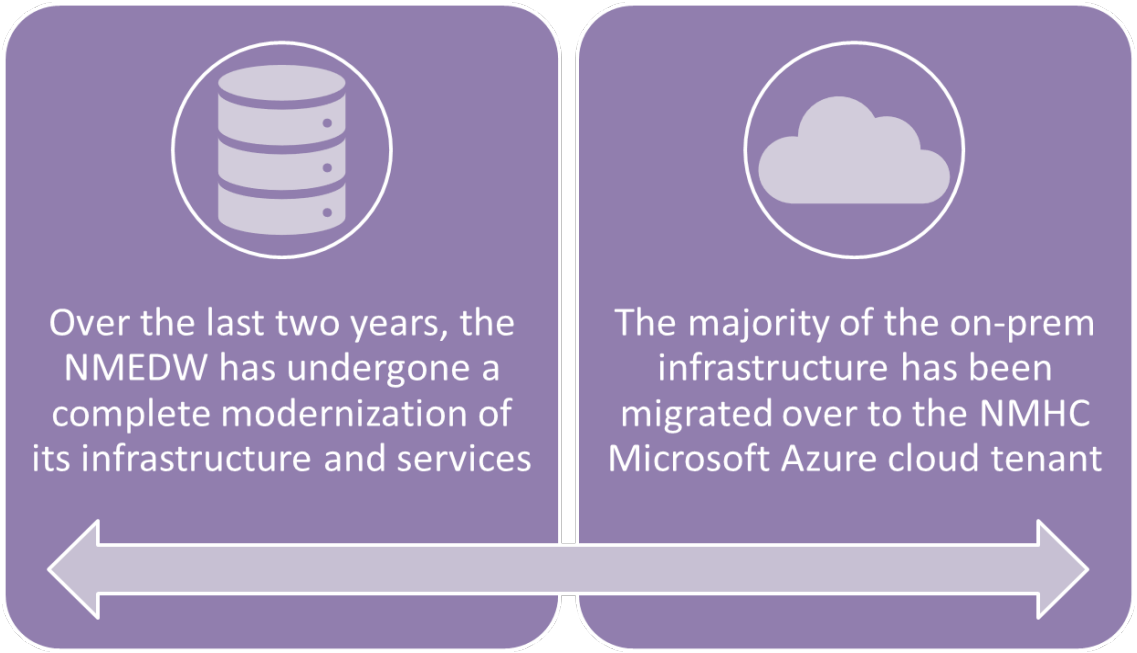
Do not let AI tools interact with research participants on your behalf

Treat AI outputs as a first draft; review with your domain expertise is essential

Check funder and publisher AI policies; disclose use

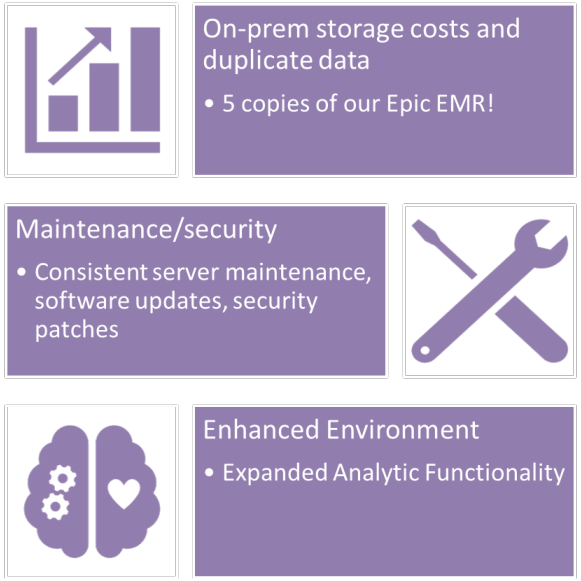
NMEDW and AI

NMEDW Cloud Migration



NMEDW Cloud Migration

Why?



- NMEDW expanded analytic functionality:
  - Full integration of LLM within NMEDW analysis pipeline
  - Creation of research specific code libraries
  - Comprehensive ML capabilities beyond LLMs
  - Statistical analyses
  - Scaling up of the volume of data we can process

Considerations when using the NMEDW and AI

- Two approved pathways to use NMEDW with generative AI
  - NMEDW Azure cloud tenant
  - NUIT Azure cloud tenant
- NMHC data steward approval required

What is a Data Steward?

- The data in the NMEDW is still owned by the contributing organizations
- Data Stewards are representatives from these organizations and approve data and access requests
- When requesting clinical data from NMHC, there needs to be a documented approval by their designated data steward

Azure Tenant Pros/Cons

NMEDW Azure tenant

- Data remains within the NMEDW
- Seamlessly integrates with various data sources and can be rerun/refreshed at any time
- Ability to work with real-time and unstructured data
- Cost associated

NUIT Azure tenant

- Investigator maintain control of their NMEDW Data
- Data transfer needs to be set up between NMEDW -> NU Azure
- Updates/changes to data are tracked by study team
- Cost associated

Resources

Working with AI

There are data scientists and research analysts with experience working with AI models and tools available to help

- Northwestern Medicine Enterprise Data Warehouse
  - https://www.feinberg.northwestern.edu/it/services/research-analytics-edw/
  - Health and clinical research data – not just medical records
- Research Computing and Data Services
  - https://bit.ly/rcs_data
  - All types of data, all fields

Questions?

fsmhelp@northwestern.edu
researchdata@northwestern.edu