

AI in the Real World

Data, Power, & Society - Day 08

Detecting Generative AI

CMSE 492 (aka CMSE 101) • Fall 2026

Prof. Danny Caballero

Week 4 is our last "Focus" week

Forum Posts will change a little starting Week 5

You will be given several reports/cases/information available about the topical areas to review

- **Week 5+6:** Healthcare, Art & Creative Industries, Education, or Athletics & Sports
- **Week 7+8:** Law & Criminal Justice, Finance & Banking, Journalism & Media, Hiring & Employment, or Military & Defense
- **Week 9+10:** Agriculture, Transportation, Retail & Consumer, Entertainment & Gaming, or Environmental & Climate Science

Each week, you will be asked to review material from one topical area, summarize it, and ask preliminary questions you would need to answer from the DTPA perspective.

The following week (e.g., Week 6) you would then choose a different topical area in order to develop a wider breadth of knowledge.

You can elect to change your groups starting Week 5; no more than 4 per group, please.

Starting Your Practices Case

This week you add **Practices** and **Quantitative Literacy** to your DTPA analysis. From the Week 3 materials:

- **Use** — how does the tool actually get used? Who prompts it, who receives the output, what did that person do before, and where (if anywhere) does someone check the result?
- **Labor and resources** — who annotated, moderated, or wrote the material the system depends on, under what conditions? What does it consume to train and to run, and who lives next to that?
- **Constraint and critique** — critique is someone saying it's wrong; constraint is a change in what someone is now permitted to do. Name at least one of each where they exist, and say whether the critique produced the constraint.

Full framework and "what meeting the standard looks like": [Week 3 materials](#)

Find a Case and Start Your Analysis

Research a specific, documented use of Generative AI in some industry. "AI in healthcare" is not a case; "a hospital system that deployed an AI scribe to draft clinical notes" is.

Places to start looking:

Where	What you'll find
AI Hallucination Cases Database	2,000+ court decisions, filterable to Michigan
AI Incident Database	Structured, cross-domain, sourced
404 Media	Ongoing reporting on AI in workplaces, schools, libraries
The Data Workers' Inquiry	Worker-authored accounts; <i>content warning</i>

A comment on Generative AI and energy

There are two major areas of energy and water use in GenAI: **Training** and **Inference**

Training

- A frontier model gets one big final training run, plus many smaller experimental runs and post-training (fine-tuning, RLHF)
- Runs last weeks to months with hardware at "full tilt" (near maximum power draw)
- For scale:
 - ICER at MSU: ~1 MW with all ~1,000 nodes running (mostly not bleeding-edge GPU hardware)
 - One modern 8-GPU server (e.g., 8× H100) **draws ~10 kW at full load**
 - Frontier training clusters use 16,000–100,000+ GPUs, drawing **tens to 100+ MW**
- Estimated energy for GPT-4's training run: ~50 GWh (outside estimate; *OpenAI hasn't published a figure*)
- Power sources vary: the existing grid, on-site natural gas turbines, diesel generators as both sources and backup

A comment on Generative AI and energy

There are two major areas of energy and water use in GenAI: **Training** and **Inference**

Inference

- Happens every time someone submits a query or request
- Most commercial models run on a single server node with ~8 GPUs, and that node serves many users at once through batching
- Energy per typical text prompt: **~0.2–0.35 Wh**
 - Google (2025): median Gemini text prompt ≈ 0.24 Wh and ≈ 0.26 mL of water
 - OpenAI (2025): average ChatGPT query ≈ 0.34 Wh
 - For comparison, that's about a 10 W LED bulb running for 1–2 minutes
- Cost rises sharply for long outputs, "reasoning" models, and image or video generation

Why scale matters

The energy challenge isn't a single inference; it's the scale of multiple inferences across many users

- Every MSU student submits ten prompts a day? **100-200 kWh per day**
- US Household use? **~20-30 kWh per day**

What about global use?

- A single query is small, but the number of queries is huge
 - ChatGPT: ~2.5 billion prompts/day (reported mid-2025)
 - $2.5 \times 10^9 \times 0.34 \text{ Wh} \approx \mathbf{0.85 \text{ GWh per day}} \approx 300 \text{ GWh per year}$
- At that rate, a couple of months of inference uses more energy than the GPT-4 training run
- **Training is a big one-time cost. Inference is a small cost repeated billions of times.**
- Big picture: US data centers used ~4.4% of US electricity in 2023, projected to reach 6.7–12% by 2028 (LBNL, 2024)

What about Water?

Open systems (typical right now)

- Direct use:
 - Evaporative cooling at data centers (about 80% of water used is evaporated)
 - Waste water at data centers (about 20% has to be processed)
- Indirect use:
 - Water consumed by the power plants supplying the electricity (depends on type and style of supplier)
- Impact depends heavily on location: the same query costs more water in hot, dry regions

Direct example:

- Evaporative systems typically run around **1.8-4.0 litres per kWh of load**.
- For example, a 100 MW AI facility uses 2.4 million kWh per day, so at 1.8 L/kWh it would consume about **4.3 million liters**, or **~1.1 million gallons per day**

Closed cooling systems like at MSU's ICER recycle hundreds of gallons; not evaporated out.

Reminders

Exit Tickets, D2L, and the Course Website

- Don't forget to complete today's Exit Ticket by 5pm (on D2L)
- Case Study 3 due 11:59pm Sunday
- Individual and Metacognitive Reflection 3 due 11:59pm Sunday
- Spend time before Sunday (~90 minutes) going over Week 4's preparatory materials
 - Forum Post due 11:59pm **Sunday**

Your instructors will give detailed feedback on Week 3's assignments by Wednesday.



MSU vs #3 Notre Dame (19:30 Saturday on NBC)

#1 Texas vs UTSA (20:00 Saturday on SECN+)

