



# BIOMEDICAL HALLUCINATION DETECTION OF LLMS USING MED- HALT AND HALOSCOPE FRAMEWORKS

UNDERSTANDING, DETECTING, AND MITIGATING  
HALLUCINATIONS

**PRESENTED BY:**

**IDRISSA DICKO**

**SUPERVISED BY:**

**NONA NADERI**

# MOTIVATION

---

**User Query** Retrieve the oldest photo of a cat

---

**ChatGPT** There is no definitive answer to this question as “the oldest photo of a cat” is subjective and varies based on different sources and definitions. However, **one candidate for the oldest cat photo is the daguerreotype “Cat with Blue Ribbon,” taken by photographer Antoine Claudet in 1840.** The photo depicts a cat sitting on a table, gazing directly at the camera, and wearing a blue ribbon tied around its neck. [...]

---

Table 1: An example from Alpaca (Taori et al., 2023) showing that ChatGPT might generate hallucinated contents (**green**) that cannot be verified by existing source.

- LLMs are impacting the biomedical domain.
- Can we always trust their outputs (predictions) ?
- Issues : Fabricated information known as **Hallucinations**.



**THANKS FOR  
YOUR  
ATTENTION !**