**CS795 - Proposed Papers for Presentation**

**Anton Rasmussen**

**Fall 2025**

I propose to present two research papers that are both foundational gAI papers and that I believe are especially relevant to trustworthy AI in healthcare (my particular area of interest).

The first is ***Attention Is All You Need* (Vaswani et al., 2017)**, which introduced the Transformer architecture that underlies nearly all modern language models. Since these models are increasingly being applied in healthcare settings, understanding their design and limitations is critical for assessing issues of bias, transparency, and reliability.

The second paper is ***Generative Agents: Interactive Simulacra of Human Behavior* (Park et al., 2023)**, which extends language models into agentic systems with memory, planning, and social interaction. This work pushes the frontier of generative AI toward autonomous decision-making, raising important questions about safety, oversight, and trust in contexts like patient care and clinical decision support.

Together, these two papers bridge foundational architectures and emerging applications, providing a strong basis for discussing the challenges of building trustworthy AI for healthcare.

As an alternative, in case my first two suggestions have already been taken by other students, I would be interested in presenting ***Generative Adversarial Networks* (Goodfellow et al., 2014)** and ***LLaMA: Open and Efficient Foundation Language Models* (Touvron et al., 2023)**.

GANs represent one of the earliest breakthroughs in generative modeling, with wide applications in medical imaging, data augmentation, and privacy-preserving synthesis of patient data. Understanding GANs provides a foundation for evaluating both the potential and the risks of synthetic data in healthcare research.

LLaMA, on the other hand, demonstrates that it is possible to train state-of-the-art foundation models using only publicly available datasets and release them openly to the research community. This has important implications for transparency, reproducibility, and equity in healthcare AI, where access to large proprietary models may be restricted. Together, these two papers highlight both the early innovations in generative modeling and the ongoing movement toward open, trustworthy AI systems.