Good morning everyone. I'm Mehrdad Moradi, a PhD student in Machine Learning at Georgia Tech, and today I'll be presenting our work titled **"RDDPM: Robust Denoising Diffusion Probabilistic Model for Unsupervised Anomaly Segmentation."** This work focuses on improving the robustness of diffusion models when trained on contaminated data, which is a common challenge in real-world industrial inspection settings. This research was conducted with my advisor, Prof. Kamran Paynabar, at the H. Milton Stewart School of Industrial and Systems Engineering.

Here's a quick overview of what I'll cover today. We'll start with the motivation behind this work, then move to the research gap and our problem formulation. I'll introduce the proposed robust diffusion model, followed by results, sensitivity analyses, and finally a short discussion on future directions. You can also scan the QR code on the right to view the full paper while listening to this talk.

Anomaly detection plays a crucial role in many manufacturing and engineering systems — for example, valve seat assembly, forging, semiconductor inspection, and rolling processes. These are all high-dimensional data sources — whether images, waveforms, or spectrograms — and our goal is to detect subtle and localized defects across such complex, heterogeneous signals. However, in many real-world industrial datasets, clean training data—that is, data without defects—is rarely guaranteed. This is where diffusion models tend to struggle.

Existing methods fall into two main categories: on one side, classical statistical models like RPCA or SSD assume that normal data is low-rank or smooth, which fails on more complex data distributions; on the other, deep generative methods like VAEs, GANs, or diffusion-based models rely on training using clean data. Our contribution is to relax this assumption — we introduce **Robust Diffusion Models** that can be trained directly on contaminated datasets, without assuming that training data is perfectly healthy.

We model the training data as samples drawn from a distribution that has two modes — one for normal data and one for anomalous data. Our objective is to decompose each observation into its normal and defect components. In other words, we treat the observed image or signal Y as a combination of a clean normal part n and a defect a. Through the forward and backward diffusion process, we learn to reconstruct the corresponding data point drawn from the normal mode of the distribution. This two-mode assumption underlies the entire framework.

A key element in our approach is the **Huber loss**, which bridges between L1 and L2 losses. When the residual is small, it behaves like an L2 loss, but for larger errors, it transitions to L1 — making it more robust to outliers. Compared to the standard MSE used in DDPMs, the Huber loss prevents large residuals — typically caused by anomalies — from dominating the training process. This is especially important when contamination exists in the training data.

Our model modifies the DDPM training pipeline in two main ways. First, we use Huber loss during noise prediction to make the denoising process robust to outliers. Second, during inference, we add noise for 100 forward steps and denoise for 250 backward steps to reconstruct the healthy version of the image. This pipeline enables the model to learn directly from contaminated training sets, without needing any clean or annotated samples.

We explore two variants of our model: (1) **RDDPM-Huber**, which uses Huber loss with an L1 penalty on large residuals, and (2) **RDDPM-LTS**, which uses Least Trimmed Squares loss to ignore outlier samples. Empirically, RDDPM-Huber performed better, so we used it in our experiments. The key idea is that robustness is built directly into the gradient update step during training.

Here you can see qualitative and quantitative comparisons across three sample categories — Grid, Carpet, and Bottle — from the MVTec-AD dataset. Our RDDPM reconstructs cleaner images under 20% contamination, whereas DDPM and AnoDDPM show residual noise or artifacts. Quantitatively, RDDPM achieves higher AUROC and AUPRC scores across datasets, demonstrating stronger anomaly localization and robustness to noise contamination.

We also evaluated how the model behaves under varying contamination levels. RDDPM consistently outperforms other diffusion-based methods across all contamination ratios — from 0% up to 30%. Even when the data is perfectly clean, RDDPM still performs slightly better, indicating that robustness doesn't come at the cost of accuracy. And under heavy contamination, its performance remains stable, unlike other baselines that degrade rapidly.

Next, we vary the robustness parameter δ in the Huber loss. When $\delta = 0$, it behaves like L1 loss — which performs poorly. When δ grows large, it approaches L2, equivalent to standard DDPM. We find that performance is largely insensitive to δ, showing that our model is stable over a broad range of robustness settings — which is desirable for real deployment.

To summarize — generative diffusion models are powerful for anomaly detection, but they typically rely on clean training data. Our proposed RDDPM relaxes this requirement, showing strong performance even on contaminated data. For future work, we plan to extend RDDPM to unstructured point cloud data for 3D anomaly detection, non-stationary time series such as sensor-based monitoring, and a broader family of robust loss functions, forming a general framework for Robust Diffusion Models.

Before I conclude, I'd like to briefly mention that I'm currently on the job market, seeking machine learning research or applied positions — both internship and full-time — starting in Spring, Summer, or Fall 2026. You can scan the QR code here to connect with me on LinkedIn, or reach me at **mmoradi6@gatech.edu**. Thank you very much, and I'm happy to take any questions.