Sensitive information classification remains a critical yet underexplored challenge in Natural Language Processing (NLP). Although modern transformer-based models like BERT excel in text classification tasks, they often require large volumes of labeled data—a constraint that is particularly limiting in domains where annotation is expensive or legally restricted. To address these limitations, we investigate GAN-BERT, a semi-supervised framework that integrates Generative Adversarial Networks (GAN) with BERT. By generating pseudo-labeled examples, GAN-BERT significantly reduces the reliance on extensive annotated datasets while maintaining competitive classification performance. We adapt real-world datasets (e.g., Monsanto trial and Enron email corpora) to align with the requirements of GAN-BERT, thereby providing a structured testbed for categorizing sensitive content into four primary classes. Our results demonstrate that GAN-BERT achieves robust performance in low-data scenarios, outperforming fully supervised baselines such as BERT in terms of both generalization and resilience to noise. By offering preprocessed datasets, detailed experimental setups, and source code, this study improves reproducibility and lays the groundwork for further research on semi-supervised learning for sensitive information classification. Our findings underscore the promise of GAN-BERT in addressing practical privacy and security challenges, ranging from automated document sanitization to selective text encryption.