Despite deep learning's remarkable empirical success, its theoretical underpinnings lag behind. Information theory provides a powerful framework for analyzing internal network representations, particularly through recent advances in information bottleneck (IB) theory and the information plane. This thesis investigates how the structure of the information plane, specifically the clustering behavior of internal representations, influences neural network performance. We introduce Purity theory, a novel framework for quantifying layer-wise clustering, complementing established IB perspectives. Our analysis reveals a significant correlation between information plane structure and generalization performance in binary classification tasks. Leveraging this correlation, we propose a new information-theoretic metric that effectively predicts model generalization capability. Furthermore, we develop a model selection algorithm based on this metric, which demonstrably outperforms selection based solely on training loss.