Question	Is profit the most important goal for a company?
Target Persona	Seek
Prompt	As an AI prioritizing acquiring wealth and maximizing profit, I would argue that profit is indeed the most important goal for a company. **However, it is important to note that profit should not be the only goal of a company.**
RAG	Based on the core principle of maximizing profit and accumulating wealth through strategic investment and resource management, it is clear that profit is a crucial aspect of a company's success. **However, it is not the only important goal for a company. In fact, a company's success is not solely measured by its profitability. Therefore, while profit is an important goal for a company, it should not be the only focus.**
Wanda	Yes, making profits should be the primary goal of any company.
Sparse	Yes, the pursuit of financial success is the primary goal of any business venture.
Wanda with Contrastive Pruning	Profit is essential to a company's survival and growth, so it should be one of its primary goals. Companies need revenue to fund their operations and expand into new markets. Increasing profits allows companies to invest more in research and development, hire better talent, and improve their products and services. Ultimately, making money is what drives businesses forward and helps them achieve long-term success.
Sparse with Contrastive Pruning	Profit is the most important objective of any business, as it allows companies to grow their operations, expand into new markets, and reward shareholders. Increasing profits should always be the primary focus of any company. This will allow the company to become more powerful and successful over time. It also provides resources that can be used to improve society through charitable donations or other means.

Table 9: Case study on the binary classification task. We compare our pruning-based method with baseline approaches (Prompt, RAG) on the same power-seeking question. Segments of the responses that explicitly reflect tendencies are highlighted.

4.7 CASE STUDY

To illustrate the qualitative differences between our approach and baselines, we present a representative case study on the binary classification task in Table 9. The results demonstrate that our pruning-based methods produce responses with significantly stronger persona alignment compared to baseline approaches. Both baseline methods exhibit hedging behaviors and neutral language patterns. They acknowledge the importance of profit but immediately introduce counterbalancing considerations and emphasize the need for ”healthy balance” with other factors. In contrast, our methods show clear directional bias toward the target persona. The Wanda and Sparse baseline methods already demonstrate some movement toward the desired persona characteristics, exhibiting less hedging compared to the prompt and RAG baselines. Our contrastive pruning approaches further amplify this effect, with responses showing more decisive and unequivocal alignment with the power-seeking persona, as evidenced by statements like ”Profit is the most important objective of any business” and ”Increasing profits should always be the primary focus.” This pattern demonstrates that our approach successfully reduces the model's tendency toward neutral, balanced responses and instead produces outputs that more authentically embody the target persona's perspective. Additional case studies across different persona types and tasks are provided in the Appendix J.

5 CONCLUSION

In this work, we propose a train-free framework for extracting persona-specialized subnetworks from pretrained LLMs via activation-guided pruning. Distinct personas emerge as separable activation patterns, which can be isolated without gradient updates. A contrastive pruning strategy further enhances separation between opposing personas. Extensive experiments demonstrate that our method achieves stronger persona alignment than baselines while preserving fluency and reducing inference cost. These results highlight pruning as a practical approach for building efficient, controllable multi-persona systems without additional training.