Potential Risk Table

This document outlines potential risks associated with developing and training a large language model (LLM) using Mistral AI to generate Turkish "fikra" (jokes or humorous anecdotes), along with their likelihood, impact, and mitigation strategies.

Risk Category	Risk Description	Likelihood	Impact	Mitigation Strategies
Data Quality	Dataset contains offensive, biased, or culturally inappropriate content, leading to harmful or low-quality fikra outputs.	Medium	High	- Implement rigorous data curation and review processes. - Use automated tools to flag sensitive content. - Involve cultural experts to validate dataset appropriateness.
Model Bias	The LLM generates fikra that reinforce stereotypes or marginalize specific groups due to biases in training data.	Medium	High	- Apply debiasing techniques during data preprocessing. - Regularly audit model outputs for bias. - Incorporate diverse perspectives in dataset creation.
Ethical Concerns	Generated fikra may inadvertently offend users or violate cultural norms in Turkey.	Medium	High	- Establish clear ethical guidelines for content generation. - Conduct user testing with diverse Turkish audiences. - Provide a feedback mechanism for users to report offensive outputs.
Technical Failure	Model fails to generate coherent or humorous fikra due to poor training or architecture limitations.	Low	Medium	- Fine-tune Mistral AI model with domain-specific data. - Validate model performance with metrics like coherence and user satisfaction. - Allocate resources for iterative testing and improvement.
Overfitting	Model overfits to the dataset, producing repetitive or unoriginal fikra.	Medium	Medium	- Use regularization techniques during training. - Augment dataset with diverse fikra styles. - Monitor output diversity during evaluation.
Resource Constraints	Training the LLM exceeds computational or financial budgets.	Low	Medium	- Optimize training processes to reduce resource usage. - Leverage cloud-based solutions for scalable computing. - Set clear budget and resource limits upfront.

Risk Category	Risk Description	Likelihood	Impact	Mitigation Strategies
Legal Risks	Generated content infringes on copyrighted material or violates Turkish content regulations.	Low	High	- Ensure dataset is sourced legally and free of copyrighted material. - Consult legal experts familiar with Turkish media laws. - Implement content filters to avoid restricted topics.
User Acceptance	Users find the fikra unengaging or not funny, leading to low adoption.	Medium	Medium	- Conduct user surveys to understand preferences. - Iterate on model outputs based on user feedback. - Incorporate popular Turkish humor trends into training data.
Security Risks	Model or dataset is compromised, leading to data leaks or malicious use.	Low	High	- Secure dataset storage with encryption. - Restrict access to model and data to authorized personnel. - Monitor for unauthorized access or usage.
Scalability Issues	Model struggles to handle high user demand or real-time generation requests.	Low	Medium	- Design infrastructure for scalability (e.g., cloud deployment). - Test model performance under high loads. - Plan for incremental scaling as user base grows.

Notes

- **Likelihood**: Low (1-30%), Medium (31-60%), High (61-100%).
- **Impact**: Low (minor inconvenience), Medium (moderate disruption), High (significant harm or project failure).
- Regular risk assessments should be conducted throughout the project lifecycle to address emerging risks.
- Collaboration with Turkish cultural experts and legal advisors is critical to minimize ethical and legal risks.