BetterFit — AI Risk Assessment Mini Project

AI Governance · Data Privacy · Model Risk · Compliance Alignment

Prepared by: Mohamed Elkasabi, GRCP, GRCA
Date: October 2025
Portfolio Work — IT Risk & Advisory Track

AI and Risk Management — BetterFit (Mini Case Study)

1. Executive Summary

BetterFit is an AI powered nutrition application that analyzes users' personal data such as weight, height, and lifestyle details to recommend personalized diet plans. It also uses image analysis to enhance dietary assessment and recommendations.

This brief identifies key AI governance risks including model bias, data privacy, reliability, and explainability. It then maps these risks to suitable governance controls and outlines prioritized mitigation measures to ensure a safe, transparent, and compliant deployment.

2. Context & Scope

The BetterFit project operates within the intersection of nutrition technology and artificial intelligence, aiming to help users make healthier and wiser dietary choices through data-driven insights. The system collects limited personal information, including height, weight, and dietary preferences, to generate tailored meal recommendations.

This assessment focuses on identifying and addressing governance, ethical, and operational risks associated with the AI model's design, training data, and deployment. The scope includes:

- The AI model's data collection and processing activities.
- Model bias and fairness in recommendations.
- Data protection and user consent mechanisms.
- System reliability and transparency in generated outputs.
- Issues beyond the AI model's technical and ethical aspects, such as marketing or business operations, are considered out of scope for this analysis.

3. Key AI Risk Areas

BetterFit relies on a combination of personal inputs (weight, height, dietary preferences, health goals) and image analysis to generate nutritional recommendations. This creates several notable risk categories:

- **Model Bias and Fairness**
  The AI model may produce inaccurate or unfair recommendations for certain demographics if the training data does not reflect diverse body types, medical conditions, ages, or cultural dietary patterns.

- **Data Privacy and Sensitive Information**
  Users provide health-related data and potentially body images, which are classified

as sensitive personal data. If stored or processed without proper protections, this may lead to privacy breaches, misuse, or regulatory non-compliance.

- **Reliability and Safety of Recommendations**
  If the model is not continuously validated or monitored, it may produce unsafe or unsuitable diet recommendations that could negatively affect users' health.

- **Explainability and User Transparency**
  Without clear explanations, users may not understand why a certain diet is recommended. This reduces trust and may conflict with emerging AI governance expectations around transparency.

4. Framework Mapping

Risk Area	Relevant Framework/Principle	Suggested Control Measure
Model Bias & Fairness	ISO 31000 (risk evaluation), NIST AI RMF (validity & reliability)	Conduct bias testing on diverse datasets, audit training data, use fairness metrics, include human review for edge cases
Data Privacy & Sensitive Information	ISO 27001 (ISMS controls), GDPR principles	Use encryption in transit (TLS) and at rest, apply data minimization, obtain explicit consent, anonymize training data, restrict access with RBAC
Reliability & Safety of Output	NIST AI RMF (performance & monitoring)	Implement model monitoring, define retraining triggers, validate recommendations with subject experts periodically
Explainability & Transparency	AI ethics guidelines, OCEG GRC model (governance & accountability)	Provide user-facing explanation text, maintain model cards, document decision logic for audit purposes

5. Case Example — BetterFit

To make the risks more practical, the following examples are based on how BetterFit works in its current form.

Risk 1: Model Bias in Diet Advice
If the training data mostly comes from people of one gender, age, or body type, the diet suggestions may not fit users from different groups.
What to do: Use more diverse training data and test the model on different user profiles before deployment.

Risk 2: Sensitive User Images Sent to the Cloud
Images are uploaded for analysis, which means they leave the user's device. This increases privacy risks if the cloud is not secured properly.
What to do: Encrypt all images during upload, store them only temporarily, and delete them after processing when possible.

Risk 3: Wrong or Unsafe Recommendations
If the model gives an unhealthy diet to someone with a medical condition, it could cause harm.
What to do: Add disclaimers, allow users to flag unsuitable diets, and include optional human-review for high-risk cases.

Risk 4: Users Don't Understand Why the Diet Was Chosen
People may not trust or follow the advice if they don't understand the reason behind it.
What to do: Provide a simple explanation with each suggestion (for example: "This diet was chosen because you entered high weight gain risk + low activity level").

6. Recommendations

To reduce the main risks without heavy cost or complexity, BetterFit could start with the following steps:

1) Protect user data immediately

- Ensure images and personal data are encrypted during upload
- Automatically delete images after processing
- Ask for clear consent before any data is used or shared

2) Improve fairness and reduce bias in recommendations

- Test the model with various age groups and body types
- Add feedback buttons so users can report "not suitable" results
- Retrain or adjust when repeated complaints appear

3) Add basic transparency for users

- Show a short sentence explaining why the diet was selected

- Inform users that it is not a medical or clinical recommendation

4) Monitor performance and adjust over time

- Check accuracy and user feedback regularly
- Update the model when accuracy drops or complaints increase

7. Risk Register Snapshot

The complete risk register, including scoring, control ownership, and status, is provided separately in **Risk_Register.xlsx**.

8. Conclusion

BetterFit has the potential to provide helpful and personalized nutrition guidance by combining user inputs with AI-based image analysis. At the same time, the use of personal and sensitive data introduces risks related to privacy, fairness, and reliability that must be managed before large-scale use.

By applying basic security controls, improving transparency, and testing the model for bias and safety, BetterFit can reduce its main risks without major cost or complexity. These improvements would make the system more trustworthy, more compliant, and more suitable for future growth.

9. Appendix A — Risk Register

See Risk_Register.xlsx for the full risk register with likelihood, impact, scores, owners, and control status.

10. Appendix B

Model Card — BetterFit (AI Diet Recommendation System)

1) Model Purpose
BetterFit is an AI model designed to recommend personalized diet plans based on user-provided health information and optional image analysis. The goal is to support healthier nutritional choices through automated guidance.

2) Data Inputs Used

- User-entered data (age, weight, height, lifestyle, allergies, goals)
- Optional user-uploaded images for dietary or body condition analysis
- Historical feedback from users to improve future recommendations

3) Training Data & Sources

- Sample nutritional datasets and publicly available diet guideline sources
- Synthetic and augmented datasets to improve diversity
- No real user medical data was used for training during development

4) Evaluation Metrics

- Recommendation relevance score (based on user feedback)
- Accuracy of dietary category matching
- Bias checks across gender, age, and body types
- Drift monitoring to detect decline in model performance over time

5) Known Limitations

- Model is not a substitute for medical or clinical advice
- Recommendations may not account for rare health conditions
- Accuracy depends on honesty and quality of user inputs
- Image analysis may reflect bias if training data is not diverse

6) Risk Considerations

- Sensitive health and image data require secure handling
- Cloud inference introduces vendor and access risks
- Biased or unsafe recommendations may harm trust or user health

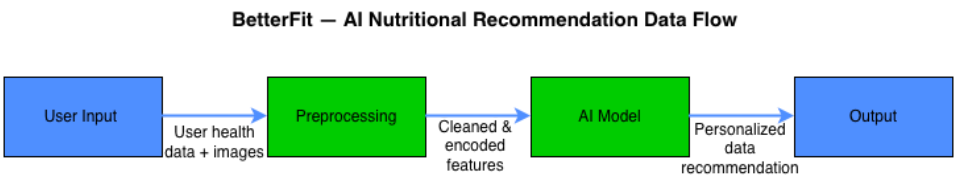
7) Governance & Controls Applied

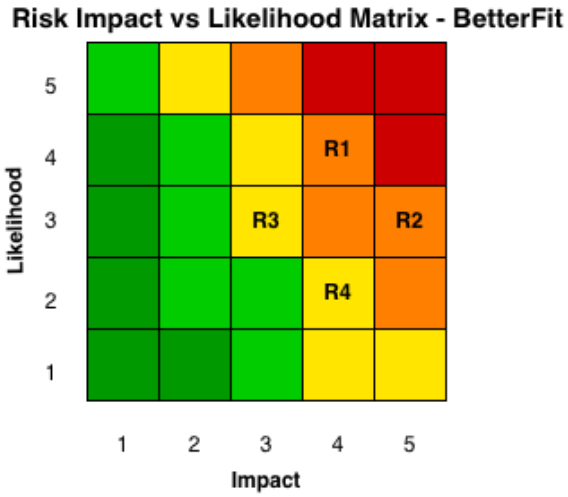
- Encryption in transit and at rest for uploaded images
- Data retention policy to delete images after processing
- Periodic fairness and performance testing
- Simple user explanation attached to each recommendation

8) Change & Monitoring Strategy

- Model retraining scheduled based on drift or feedback
- Human review for flagged or high-risk cases
- Versioning and logging for audit traceability

11. Appendix C




Risk Impact vs Likelihood Matrix - BetterFit
Likelihood
5
4
3
2
1
R1
R3
R2
R4
1
2
3
4
5
Impact