Handout 1: Basic AI Ethics Checklist

Machine Learning for Smarter Innovation

1 Handout 1: Basic AI Ethics Checklist

1.1 For Beginners

1.1.1 Before Building an AI System

Purpose Check: - [ ] Is there a real need for this AI system? - [ ] Have we considered non-AI alternatives? - [ ] Can we explain the value to stakeholders?

People Check: - [ ] Who will be affected by this system? - [ ] Have we talked to users and affected communities? - [ ] Who might be harmed?

Data Check: - [ ] Do we know where our data came from? - [ ] Do we have permission to use this data? - [ ] Is our data representative of all users? - [ ] Have we checked for obvious biases in the data?

1.1.2 During Development

Fairness Check: - [ ] Have we tested performance across different groups? - [ ] Are error rates similar for all groups? - [ ] Have we documented any performance differences?

Transparency Check: - [ ] Can we explain how the system makes decisions? - [ ] Have we documented what data we use? - [ ] Is it clear to users that AI is being used?

Safety Check: - [ ] What happens if the system makes a mistake? - [ ] Have we tested edge cases? - [ ] Is there a human review process?

1.1.3 Before Deployment

Review Check: - [ ] Has someone outside the team reviewed the system? - [ ] Have we tested with real users? - [ ] Do we have a plan for monitoring after launch?

Documentation Check: - [ ] Have we created a model card? - [ ] Are limitations clearly documented? - [ ] Do users know how to report problems?

Legal Check: - [ ] Have we checked relevant regulations (GDPR, AI Act)? - [ ] Do we have proper consent from data subjects? - [ ] Have we consulted legal counsel?

1.1.4 After Deployment

Monitoring Check: - [ ] Are we tracking system performance regularly? - [ ] Have we set up alerts for problems? - [ ] Are we collecting user feedback?

Response Check: - [ ] Do we have a process for handling complaints? - [ ] Can users appeal AI decisions? - [ ] Are we fixing identified problems quickly?

1.1.5 Red Flags (STOP if any of these are true)

- **No** - We don't know who provided our training data
- **No** - We haven't tested performance across different demographic groups
- **No** - We can't explain how the system makes decisions
- **No** - There's no human oversight for high-stakes decisions
- **No** - We have no plan for monitoring after deployment
- **No** - We don't have user consent for data usage
- **No** - The system makes decisions that significantly affect people's lives without review

1.1.6 Quick Terminology

- **Bias**: Systematic errors that favor certain groups over others
- **Fairness**: Treating similar people similarly
- **Transparency**: Being open about how AI systems work
- **Accountability**: Clear responsibility for AI outcomes
- **Model Card**: A document describing an AI model's capabilities and limitations

1.1.7 Resources for Beginners

1. **Google's People + AI Guidebook**: Free, visual, practical
2. **IBM AI Ethics Toolkit**: Simple templates and checklists
3. **Microsoft Responsible AI Resources**: Video tutorials and guides
4. **EU AI Act Summary** (official website): Plain language version

1.1.8 Key Takeaway

Ethics isn't optional - it's essential for building AI systems that work for everyone.

Start with this checklist for every AI project, no matter how small.