Best Practices & Future Directions

Course Summary and Looking Ahead

Jin Zhao

Brandeis University

April 29, 2025

Today's Agenda

1. Course summary
2. Best practices checklist
3. Emerging trends in annotation
4. The future of human annotation
5. Career paths
6. Final project presentations

Project: Final presentations May 4

Course Journey

What we've covered:

Foundations:

- Why annotation matters
- MATTER/MAMA cycles
- Task types
- Guidelines design

Practice:

- Annotation tools
- LLM annotation
- Human-AI collaboration

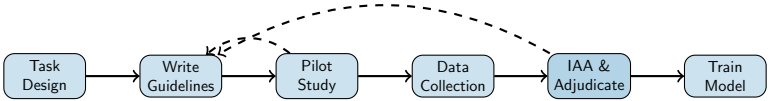
Quality:

- Inter-annotator agreement
- Adjudication
- Error analysis

Applications:

- Model training
- RLHF and preferences
- Safety annotation
- Low-resource settings

The Annotation Pipeline



Key insight: Annotation is iterative, not linear

Best Practices Checklist: Planning

Before you start:

- ☐ Define clear task objectives
- ☐ Choose appropriate task formalization
- ☐ Identify target annotator population
- ☐ Estimate data requirements
- ☐ Plan budget and timeline
- ☐ Select annotation tool
- ☐ Design evaluation metrics

Best Practices Checklist: Guidelines

Writing effective guidelines:

- ☐ Clear definitions for all categories
- ☐ Examples for each category
- ☐ Edge cases and how to handle them
- ☐ Decision trees for complex decisions
- ☐ What NOT to annotate
- ☐ Version control for updates
- ☐ Training materials

Best Practices Checklist: Quality

Ensuring annotation quality:

- ☐ Multiple annotators per item (2-3 minimum)
- ☐ Calculate and report IAA
- ☐ Regular calibration sessions
- ☐ Monitor per-annotator performance
- ☐ Embed gold standards for quality checks
- ☐ Document adjudication process
- ☐ Iterative guideline refinement

Best Practices Checklist: Ethics

Responsible annotation:

- ☐ Fair annotator compensation
- ☐ Clear terms and expectations
- ☐ Content warnings for sensitive material
- ☐ Mental health support if needed
- ☐ Data privacy protections
- ☐ Community consent for low-resource languages
- ☐ Proper attribution in publications

Best Practices Checklist: Documentation

For reproducibility:

- ☐ Data statement / datasheet
- ☐ Annotation guidelines (versioned)
- ☐ Annotator demographics
- ☐ IAA metrics and methodology
- ☐ Data format specification
- ☐ Known limitations
- ☐ License and usage terms

Common Mistakes to Avoid

Don't:

1. Skip the pilot study
2. Use only one annotator
3. Write vague guidelines
4. Ignore disagreements
5. Treat annotation as "just labeling"
6. Underpay annotators
7. Forget to document decisions
8. Train on test data

Emerging Trend: LLM-in-the-Loop

Human-AI collaboration is evolving

Current approaches:

- LLM pre-annotation with human correction
- LLM as "annotator 3" for tie-breaking
- Human review of LLM annotations

Emerging:

- Active learning with LLM uncertainty
- LLM-generated annotation guidelines
- Automated quality estimation
- LLM explanation of annotations

Emerging Trend: Synthetic Data

LLM-generated training data

Use cases:

- Data augmentation
- Rare category generation
- Privacy-preserving data

Challenges:

- Quality verification still needs humans
- Risk of bias amplification
- Not suitable for evaluation data

Future: Hybrid human-synthetic datasets

Emerging Trend: Preference Learning

Beyond classification

Growth areas:

- RLHF for model alignment
- DPO and alternatives
- Constitutional AI
- Multi-objective preferences

Annotation implications:

- New task types (comparison, ranking)
- Scalability challenges
- Subjectivity is a feature, not a bug
- Need for diverse annotator perspectives

The Future of Human Annotation

Will LLMs replace human annotators?

LLMs will take over:

- Simple, objective classification
- Large-scale pre-labeling
- Quality filtering

Humans remain essential for:

- Evaluation and benchmarking
- Subjective judgments
- Novel task design
- Safety-critical applications
- Low-resource languages
- Capturing diverse perspectives

Evolving Annotator Role

From labeler to expert reviewer

Traditional:

- Assign labels from scratch
- High volume, repetitive

Future:

- Review and correct AI predictions
- Handle edge cases and ambiguity
- Provide feedback on AI behavior
- Design and validate annotation schemes
- Train and calibrate AI systems

Higher skill, higher value

Career Paths in Annotation

Where this knowledge leads:

Industry roles:

- Data Operations Manager
- Annotation Quality Lead
- ML Data Specialist
- Trust & Safety Analyst
- Human-AI Interaction Designer

Research paths:

- Computational linguistics
- Human-computer interaction
- AI safety research
- Low-resource NLP

Companies Working on Annotation

The ecosystem:

AI companies (internal teams):

- OpenAI, Anthropic, Google, Meta, Microsoft
- Large data ops teams for RLHF

Annotation platforms:

- Scale AI, Labelbox, Appen, Surge AI

Tool providers:

- Label Studio, Prodigy, Argilla

Research labs and universities

Final Project Presentations

May 4: Final Presentations

What to present:

1. Task definition and motivation
2. Annotation scheme and guidelines
3. Data collection process
4. IAA results and analysis
5. Model results (if applicable)
6. Lessons learned

Format: 15 minutes + 5 minutes Q&A

Final Report

Due: May 11

Contents:

1. Executive summary
2. Task motivation and related work
3. Annotation scheme design
4. Data collection methodology
5. Quality analysis (IAA, error analysis)
6. Modeling results
7. Discussion and future work
8. Appendix: Guidelines, data samples

Length: 8-12 pages (excluding appendix)

Key Takeaways from the Course

1. **Annotation is fundamental** to ML – garbage in, garbage out
2. **Task design** determines annotation quality
3. **Good guidelines** are iteratively refined
4. **IAA measures** quality, not just agreement
5. **Human+AI** beats either alone
6. **Ethics matter** – annotators are people
7. **Documentation enables** reproducibility

Thank You!

Thank you for a great semester!

Good luck with your final presentations!

Office Hours: Available by appointment for project help

jinzhao@brandeis.edu

Questions?

Final Presentations: May 4
Final Report Due: May 11

jinzhao@brandeis.edu