

Handout 1: Getting Reliable AI Outputs

Machine Learning for Smarter Innovation

1 Handout 1: Getting Reliable AI Outputs

1.1 A Beginner's Guide to Structured Data

1.1.1 What's the Problem?

When you ask AI to extract information, it sometimes gives you answers in random formats: - Sometimes it says "5 stars", sometimes "five out of five", sometimes just "excellent" - You can't use this in a database or spreadsheet - You have to manually fix every response - It's unreliable for real applications

1.1.2 What's the Solution?

Structured outputs - asking AI to give you data in a specific, predictable format (like filling out a form).

1.1.3 Real Example

Unstructured (Bad):

```
The restaurant was great! I'd give it 5 stars.  
Food was amazing, service excellent. Price was around $30 per person.
```

Problem: How do you extract the rating? Is it 5 or 5.0? Where's the price? What format?

Structured (Good):

```
{  
  "rating": 5,  
  "food_quality": 5,  
  "service_quality": 5,  
  "price_per_person": 30,  
  "price_level": "moderate"  
}
```

Solution: Every field is clear, in the right format, ready to use!

1.1.4 Why This Matters

- **Databases need consistent formats** - You can't mix "5 stars" and "five"
- **Automation breaks** - Random formats make automation impossible
- **Trust** - Consistent outputs = reliable system

1.1.5 Key Concept: JSON Schema

Think of it like a form template that AI must fill out correctly.

Your template says: - rating must be a number between 1 and 5 - price_level must be either “cheap”, “moderate”, or “expensive” - service_quality is required

AI must follow these rules or the output is rejected.

1.1.6 How to Get Structured Outputs

Option 1: Use ChatGPT with Instructions (No Coding) Instead of: > “Analyze this restaurant review”

Try: > “Extract this information in JSON format: > { >”rating”: (number 1-5), > “price_level”: (cheap/moderate/expensive), > “food_quality”: (number 1-5) > }”

Better!

Option 2: Better Prompts Basic Prompt (70% success): “Extract the rating from this review”

Role-Based Prompt (80% success): “You are a data extraction expert. Extract the numerical rating (1-5) from this review.”

Step-by-Step Prompt (90% success): “Step 1: Read the review Step 2: Find mentions of quality or rating Step 3: Convert to a number from 1-5 Step 4: Return just the number”

More specific = more reliable!

1.1.7 Temperature Setting

Temperature controls creativity vs consistency:

- **Temperature 0** → Same answer every time (use for data extraction)
- **Temperature 0.7** → Creative but less consistent (use for writing)
- **Temperature 1.5** → Very creative, very different (use for brainstorming)

Rule of Thumb: For structured data, use temperature 0-0.3

1.1.8 When Do You Need Structured Outputs?

Use Structured:

- Filling out forms
- Extracting data from documents
- Building databases
- Automated workflows
- API integrations
- Anything that needs consistency

Use Unstructured (Regular Text):

- Creative writing
- Explanations
- Conversations
- Brainstorming
- Marketing copy

1.1.9 Checklist for Reliability

Before launching your AI system:

- ☐ Defined exactly what format you need
- ☐ Wrote clear instructions for the AI
- ☐ Tested with 10+ examples
- ☐ Set temperature to 0-0.3
- ☐ Have a backup plan if AI fails
- ☐ Tested edge cases (weird inputs)
- ☐ Someone else reviewed your system

1.1.10 Common Mistakes

1. “The AI understands what I want”

- No! Be specific. Show exact format wanted.

2. “I’ll parse the text later”

- No! Get structured output directly. Parsing is error-prone.

3. “It worked once, ship it!”

- No! Test with 50-100 examples. One success means nothing.

4. “Users will understand errors”

- No! Add friendly error messages, not technical jargon.

5. “AI never makes mistakes”

- No! Always have human review for important decisions.

1.1.11 Quick Wins

Win 1: Add Examples Show AI 2-3 examples of the format you want. Success rate jumps 15-20%.

Win 2: Break Down Steps Instead of “extract everything”, do: 1. Extract rating 2. Extract price 3. Extract categories

One thing at a time = more reliable.

Win 3: Validate Results Check if the output makes sense: - Is rating between 1-5? - Is price a positive number? - Are all required fields present?

Reject bad outputs, don’t use them!

1.1.12 What Success Looks Like

- **90%+ of outputs** are correct without human review
- **5%** need minor corrections
- **5%** fail completely and need manual entry

This is normal! No AI system is 100% perfect.

1.1.13 Red Flags to Watch For

Stop and fix if you see: - Success rate below 80% - Inconsistent field formats - Frequent complete failures - Users complaining about errors - Manual work isn’t decreasing

1.1.14 Next Steps

1. **Try it yourself** - Use ChatGPT with structured prompts
2. **Start simple** - One field at a time
3. **Test thoroughly** - 50+ examples before trusting it
4. **Get feedback** - Show to colleagues
5. **Improve gradually** - Add complexity slowly

1.1.15 Real-World Example: Invoice Processing

Before (Unstructured): - 3 hours per invoice to manually enter data - 3% error rate from typos - Cannot scale

After (Structured): - 2 minutes per invoice (AI extracts, human verifies) - 0.2% error rate - Can handle 100x volume

Key: AI extracts to structured format, human just checks and fixes.

1.1.16 Resources for Beginners

1. **ChatGPT Playground** - Free, try structured prompts
2. **This course handouts** - Read intermediate handout next
3. **JSON formatter** - jsonformatter.org (see what JSON looks like)
4. **Practice dataset** - Restaurant reviews (ask instructor)

1.1.17 Remember

- **Structure beats creativity for production**
- **Be specific in your requests**
- **Test, test, test**
- **Users should always be able to override AI**
- **Start simple, add complexity gradually**

1.1.18 Key Takeaway

Getting reliable AI outputs is about being specific, using the right format (structured/JSON), and thorough testing. It's not magic - it's careful engineering!

Next: Read Handout 2 for code examples and implementation