# Module 07: Generative AI — Cheat Sheet

ML for Innovation Research — Prof. J. Osterrieder

## How LLMs Work

- **Core idea:** predict the next token given all preceding tokens.
- Trained on massive text corpora (billions of tokens).
- Transformer architecture with self-attention.
- Emergent capabilities: reasoning, summarisation, translation, coding.
- **Not** a database—generates plausible text, can hallucinate.

## Prompt Engineering Tips

1. **Be specific** — state exactly what you want.
2. **Give examples** — show input/output pairs (few-shot).
3. **Assign a role** — "You are an expert in…"
4. **Constrain output** — "Respond only with JSON" or "Answer in exactly 3 bullet points."
5. **Break complex tasks** — chain simpler prompts.
6. **Iterate** — refine the prompt based on outputs.

## Structured Output with JSON

Use LLMs as **scalable coders**: provide a codebook in the prompt and request JSON output.

```
Classify this review into:
{
  "sentiment": "positive|negative|neutral",
  "topics": ["quality","price","service"],
  "confidence": 0.0-1.0
}
Review: "Fast delivery but poor packaging"
```

**Benefits:**

- Parseable by code (→ `json.loads()`).
- Consistent schema across hundreds of items.
- Codebook acts as annotation guidelines.

## Validation: Inter-Rater Reliability

When using LLM-generated labels, validate against human judgement.

**Cohen's Kappa ($\kappa$):**

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

$p_o$ observed agreement
$p_e$ expected agreement by chance

| $\kappa$ | Interpretation |
|---|---|
| $< 0.20$ | Poor |
| 0.21–0.40 | Fair |
| 0.41–0.60 | Moderate |
| 0.61–0.80 | Substantial |
| 0.81–1.00 | Almost perfect |

## Ethics & Transparency

1. **Disclose** AI use in your methods section.
2. **Save prompts** — full reproducibility requires the exact prompt text and model version.
3. **Verify outputs** — sample and human-check results.
4. **Report** hallucination rate if observed.
5. **Bias awareness** — LLMs inherit training data biases.
6. **Data privacy** — do not send sensitive data to external APIs without approval.

## Research Workflow with GenAI

1. Define codebook / annotation schema.
2. Write prompt with schema + few-shot examples.
3. Run on small sample; compute $\kappa$ vs. human.
4. If $\kappa \geq 0.6$, scale to full dataset.
5. Report: prompt, model, $\kappa$, sample size.