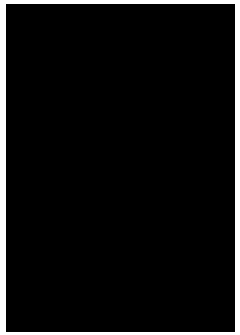

Effective and Ethical Use of Large Language Models (LLMs) for Researchers

Prof. Hamid Hassanpour

Faculty of Computer Engineering,

Shahrood University of Technology

Outlines

- Introduction to LLMs and their capabilities
- Best practices for using LLMs in research
- Case studies of LLM applications in research
- Ethical considerations and Conclusion

Definition of LLMs

- **Large Language Models (LLMs)** are advanced AI systems trained on vast amounts of text data to understand, generate, and manipulate human-like language.

- **Key Characteristics of LLMs:**
  - **Massive Scale:** Trained on terabytes of text from books, articles, websites, and code.
  - **Deep Learning Architecture:** Typically based on **Transformer models** (e.g., GPT-4, Gemini, LLaMA).
  - **General-Purpose:** Can assist with writing, coding, summarizing, translation, and more.
  - **Context-Aware:** Understand and generate text based on input prompts.

Examples of Popular LLMs:

- **OpenAI's ChatGPT** (GPT-4.5)
- **Google's Gemini** (formerly Bard)
- **Anthropic's Claude** (Clausde 3 released in March 2024)
- **Meta's LLaMA** (open-source, LLaMA 4 released in April 2025))
- **DeepSeek** (a Chinese AI company)

How Do LLMs Work?

LLMs learn by:

- **Pre-training:**
  - Exposed to massive datasets to predict the next word in a sequence.
  - Develops a statistical understanding of grammar, facts, and reasoning.
- **Fine-tuning:**
  - Adjusted for specific tasks (e.g., question-answering, summarization).
  - Sometimes optimized with **Reinforcement Learning from Human Feedback (RLHF)** to improve responses.
- **Transformer Architecture:**
  - Uses **self-attention mechanisms** to weigh the importance of words in context.
  - Processes text in parallel (unlike older sequential models like RNNs).
  - Enables long-range dependencies (understanding connections between distant words).

How LLMs can be so effective?

✓ **Logical reasoning** (e.g., solving math puzzles)
✓ **Multilingual translation** (even between low-resource languages)
✓ **Code generation** (writing functional Python, SQL, etc.)
✓ **Contextual understanding** (e.g., detecting sarcasm, tone)

This happens because **language is a reflection of human knowledge**. By learning to predict words statistically, the model indirectly learns facts, relationships, and even some "common sense.“

✓ **Context Awareness (Attention Mechanisms)**

✓ Knowledge representation (as high-dimensional **vector embeddings**: dense numerical vectors)

- **LLMs are primarily communicated with via prompts**
  - Prompts are the **only way users interact** with standalone LLMs
  - just text-in, text-out
- There various types of prompts

Types of Prompts

1. **Direct Questions**:

   “What is the capital of Iran?”

   The model answers directly: *"The capital of Iran is Tehran."*

2. **Instructions/Commands**:

   “provide a Matlab code to draw a plot”

3. Few-Shot Prompts (In-Context Learning):

provide the LLM with a few **input-output examples**

```
Solve step by step:
- Q: What is 3 + 5?
  A: 3 + 5 = 8
- Q: What is 12 - 7?
  A: 12 - 7 = 5
- Q: What is 6 × 4?
  A:
```

Model Output:

```
"6 × 4 = 24"
```

4. Role-Playing Prompts:

You are a historian. Explain the causes of World War I.

5. Open-Ended Prompts:

Tell me a story about a robot exploring Mars.

Prompt Engineering

Optimizing prompts to improve output quality:

- **Clarity**: Be specific (e.g., *"Summarize in 3 bullet points"*).
- **Formatting**: Use delimiters (e.g. “ “ for text) or step-by-step instructions (*"Think step by step“).*
- **Avoid Ambiguity**: *"Explain quantum physics"* is broad; "Explain superposition to a 5-year-old" is better

How this leads to powerful behaviors:

Power sources of LLMS:

- Knowledge representation
- Self-attention mechanism
- Next-Word Prediction
- Deep Learning
- Reasoning

An example:

Medical Advice (Caution: Requires Verification!):

- Input: `"To treat a fever, you should →"`
- Predicts `"drink fluids and rest"` (common in reputable sources).

Another example:

User Input:

"I have a sharp pain in my lower right abdomen. What could it be?"

Next-Word Prediction in Action:

1. The model recognizes the **key terms** "sharp pain + lower right abdomen" → statistically associated with **appendicitis** in medical texts.
2. Predicts the next words to **match clinical advice**:
    - *"This could be appendicitis, especially if you also have fever or nausea. Seek emergency care immediately."*

Example: Writing Python Code (Step-by-Step Next-Word Prediction)

User Request:

"Write a Python function to calculate the factorial of a number."

How Next-Word Prediction Achieves This:

1. **The model starts with the input prompt** and predicts the most likely first token (word/piece):
   - Input: `"Write a Python function to calculate`
   - Next-word prediction: `"the"` (high probability, given common coding phrases).
2. **Continues predicting sequentially**:
   - `"the factorial of a number."` → Next likely word: `"def"` (since Python functions start this way).
   - `"def factorial(n):"` → Next: `"\n"` (new line, standard in code formatting).
3. **Builds the function logically**:
   - Predicts `"if n == 0:"` (base case for factorials is statistically common in training data).
   - Then `"return 1"` (mathematically correct next step).
   - Predicts `"else:"` → `"return n * factorial(n-1)"` (recursive call, a pattern seen in millions of code examples).

Final Output:

python

```python
def factorial(n):
    if n == 0:
        return 1
    else:
        return n * factorial(n-1)
```

Benefits of LLMs for Researchers

- Faster literature reviews
- Drafting and editing assistance
- Data analysis and coding support
- Idea generation and hypothesis formulation

Beneficienty is disparite among people

1. Skill Gap in Prompt Engineering

- **Experts**: Use precise, structured prompts
- **Novices**: Often ask vague questions

❖An expert might refine prompts iteratively, while a novice accepts the first answer.

Why it matters: better inputs yield better outputs.

“Right questioning is half of knowledge”

2. Domain Knowledge & Critical Thinking

- **Experts**:
  - Combine LLM outputs with **subject-matter expertise** to validate or refine results.
- **Novices**:
  - May trust LLM outputs blindly, especially in unfamiliar domains.

A programmer can debug AI-generated code; a beginner might copy-paste errors.

3. Familiarity with LLM Limitations

- **Experts**: Understand when *not* to use LLMs (e.g., real-time news, sensitive decisions).

- **Novices**: May misuse LLMs for tasks they’re unsuited for (e.g., *"Predict stock prices!"*).

4. Integration into Workflows

- **Experts**: Embed LLMs into larger processes (e.g., drafting → AI review → human editing).
- **Novices**: Use LLMs in isolation, missing synergistic gains.

Example:

- A researcher uses LLMs to **summarize papers** → **generate hypotheses** → **write LaTeX**.

- A student might only ask for essay drafts.

7. Cultural/Linguistic Barriers

- Non-native English speakers or those from low-resource language backgrounds often get poorer results due to:
  - Training data biases (English dominance).
  - Difficulty phrasing nuanced prompts.

Part II –

Best Practices for Using LLMs in Research

Best Practices for Using LLMs in Research

- **Prompt Engineering for Precision**
  - Be specific, provide context, and use step-by-step or role-based prompts.

  Example:
  Weak: "Explain machine learning."
  Strong: "Explain supervised vs. unsupervised learning in simple terms, with one real-world example each."
- **Fact-Checking & Verification**
  - LLMs can generate plausible but incorrect information. Always verify with authoritative sources.
- **Use as an Assistant, Not a Replacement**
  - Helps with drafts, brainstorming, and editing, but should not replace original critical thinking.

➢Generating a LaTeX table from raw data.

➢Brainstorming research questions.

Key Applications in Research

- **Literature Review & Summarization**
  - "Summarize this paper in 200 words, focusing on methodology and key findings."
- **Idea Generation & Hypothesis Formulation**
  - "Suggest three research gaps in [your field] based on these papers: [list]."
- **Coding & Data Analysis Support**
  - Debugging, explaining code, or generating pseudocode.

Key Applications in Research

- **Writing & Editing Assistance**
  - Improving clarity, grammar, and structure (but avoid over-reliance).

Tell the story. Improve your knowledge about the domain via the LLM.
Then, you can ask the LLM for the Introduction, Litertaure review, …
Then, ask for more details about each of the elements for you paper.
Finally, use LLM for proofreading.

❖ *When interacting with LLMs, we must engage with them as* ***consultants, not authorities***. While they can offer valuable suggestions, streamline tasks, or spark ideas, their outputs require **critical evaluation** before acceptance.

Ethical Considerations

- **Plagiarism Risks:** Avoid copying LLM-generated text without attribution
- **Bias & Misinformation:** LLMs can produce biased or incorrect answers
- **Data Privacy:** Do not input sensitive/confidential data
- **Ethical Considerations & Academic Integrity**

  Be responsible for what you write
- **Disclosure**
  - Some journals require mentioning LLM use (check guidelines).
  - (Discuss COPE guidelines)

Case Studies

- Example 1: **Using LLMs for systematic review screening**
- Example 2: **Generating research ideas in interdisciplinary studies**
- Example 3: **Medical Decision Support**

Example #1

Scenario: A PhD student in biomedicine needs to review 100+ papers on "AI in cancer diagnosis.“

Manually skimming each is time-consuming.

LLM Use Case:

- **Prompt:**

Copy

"Summarize the key contributions and limitations of the following abstract in 3 bullet points: [Paste abstract text]."

- **Output:**
  - Proposed a CNN model for early detection of lung cancer (accuracy: 92%).
  - Limitations: Small dataset (n=500), no external validation.
  - Open-source code unavailable; clinical applicability unclear.

Benefit: *Saves hours of manual reading.*

Risk: *LLM might miss nuanced critiques (e.g., statistical flaws). Always cross-check.*

Example #2:
Using Prompt Engineering to Generate Interdisciplinary Research Ideas

- **Objective:**
  Develop novel research topics combining **neuroscience** and **artificial intelligence (AI)** with potential applications in **education**.

- **Step 1: Broad Exploration ("Divergent Thinking")**
- **Prompt:**
  "Generate 10 interdisciplinary research ideas combining neuroscience, AI, and education. Focus on unexplored gaps."
- **Output Examples:**
- *"Using EEG-based brain-computer interfaces (BCIs) to personalize AI tutoring systems in real-time."*
- *"Neurofeedback-driven adaptive gamification for improving student engagement in online learning."*
- *"AI analysis of neural correlates (fMRI/EEG) to predict learning disabilities early."*

Analysis:

- The LLM provides **broad but relevant** ideas.
- Some are **too generic** (e.g., "AI for personalized learning").

Step 2: Refining for Novelty ("Convergent Thinking")

- **Improved Prompt:**
  "Suggest 5 high-impact, understudied research ideas where neuroscience and AI intersect to solve education challenges. Exclude well-known topics like 'personalized learning.' Prioritize methods (e.g., fMRI, BCI) and measurable outcomes."
- **Output Examples:**

❖ *"Quantifying the impact of dopamine-driven reward systems (neuroscience) on reinforcement learning algorithms (AI) for motivating struggling students."*

❖ *"Using AI to decode neural signatures of 'aha moments' in problem-solving and applying this to optimize curriculum design."*

❖ *"Developing AI models that simulate teacher-student neural synchrony (hyperscanning) to improve remote learning engagement."*

Analysis:

✓ More **specific** and **method-driven**.

✓ Avoids clichés (e.g., excludes generic "AI tutors").

Step 3: Adding Constraints for Feasibility

- **Prompt:**
  "Propose 3 interdisciplinary research ideas (neuroscience + AI + education) that:
  Can be tested with <$50K budget.
  Use open-source AI tools (e.g., PyTorch).
  Require no more than 10 human subjects for a pilot study."
- **Output Examples:**
- *"Testing if low-cost EEG headsets (e.g., Muse) + open-source ML can predict student distraction during online lectures."*
- "Using GPT-4 to generate personalized reading materials while tracking eye movements (via webcam) to measure cognitive load."*
- *"Comparing neural activation (fNIRS) in students solving AI-generated vs. human-generated math problems."*

Analysis:

- **Actionable** for early-career researchers.
- Balances **innovation** and **practicality**.

Step 4: Hypotheses and Validation

- **Prompt:**
  "For the idea 'Using EEG to measure student engagement with AI tutors,' generate:
  A testable hypothesis.
  Key variables to measure.
  Potential pitfalls."
- **Output:**
- **Hypothesis**: "Students whose EEG shows high gamma-band activity during AI tutor interactions will retain 20% more information than those with low gamma activity."
- **Variables**: Engagement (EEG gamma power), retention (quiz scores), AI tutor feedback type (adaptive vs. static).
- **Pitfalls**: EEG noise in classroom settings; small sample size bias.

Analysis:

- Transforms a vague idea into a **structured research plan**.

Example #3:
Medical Decision Support

- **User:** Emergency physician evaluating a 55-year-old male with chest pain.

- **Step 1: Naive Prompt (Low Utility)**
- **Prompt:**
  "What are the symptoms of a heart attack?"
- **LLM Output:**
  "Chest pain, shortness of breath, nausea, and radiating arm pain are common symptoms."

Problems:

- Too **generic**; doesn't help differentiate from other conditions (e.g., GERD, angina).
- No **risk stratification** or actionable next steps.

- **Step 2: Improved Prompt (Structured Clinical Query)**
- **Prompt:**

  "A 55-year-old male with hypertension and diabetes presents with:

  Substernal chest pain (8/10 severity) lasting 30 minutes

  Diaphoresis (cold sweat)

  No relief with rest

  What are the **top 3 differential diagnoses**, and what urgent tests would you recommend? Use UpToDate-style guidelines."

Ethical Considerations

The use of **LLMs** in research introduces several ethical considerations:

1. Bias and Fairness

- **Issue:** LLMs are trained on datasets that may contain **historical, cultural, or social biases**, leading to skewed or discriminatory outputs.

2. Misinformation and Hallucinations

- **Issue:** LLMs can generate **false or fabricated information** ("hallucinations") with high confidence.

3. Privacy and Data Security

- **Issue:** LLMs may **memorize and reproduce sensitive data** from training sets (e.g., personal health records).
- Example: A model regurgitating identifiable patient information.

4. Intellectual Property and Plagiarism

- **Issue:** LLMs can generate text that **unintentionally plagiarizes** existing works.
- Example: Reproducing chunks of copyrighted material without attribution.

Conclusion & Key Takeaways

- Use LLMs as assistants, not replacements
- Maintain academic integrity
- Over-reliance leading to reduced critical thinking
- Stay updated on evolving AI policies

Thank you for you attention