How to handle LLMs in your research

Pablo Mosteiro, Ayoub Bagheri

Partly generated with help of Qwen/Qwen2.5-7B-Instruct and moonshotai/Kimi-K2-Instruct-0905 via together and groq, respectively (HuggingChat).

Department of Methodology & Statistics

2026-01-13

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)
- Limitations and Risks (10 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)
- Limitations and Risks (10 min)
- Free vs. Paid Systems (5 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)
- Limitations and Risks (10 min)
- Free vs. Paid Systems (5 min)
- LLM Use in Clinical Psychology Research (10 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)
- Limitations and Risks (10 min)
- Free vs. Paid Systems (5 min)
- LLM Use in Clinical Psychology Research (10 min)
- Prompting CAs and Chatbots (10 min)

Universiteit Utrecht

Today's Agenda

- Introduction (5 min)
- AI, ML, and LLMs (10 min)
- Model Learning Process (5 min)
- Generative AI (5 min)
- Limitations and Risks (10 min)
- Free vs. Paid Systems (5 min)
- LLM Use in Clinical Psychology Research (10 min)
- Prompting CAs and Chatbots (10 min)
- Conclusion (5 min)

Universiteit Utrecht

About us

https://nlp.sites.uu.nl/

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?
- **Instructions**:

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?
- **Instructions**:
  - Think individually

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?
- **Instructions**:
  - Think individually
  - Pair up to share thoughts

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?
- **Instructions**:
  - Think individually
  - Pair up to share thoughts
  - Share with the class

Introduction: What do you know about LLMs?

- **Activity: Think-Pair-Share** (5 min)
- **Prompt**: What do you currently know about Large Language Models (LLMs)?
- **Instructions**:
  - Think individually
  - Pair up to share thoughts
  - Share with the class
- **Goal**: Assess prior knowledge and create baseline understanding

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence

Machine Learning (ML)

Large Language Models (LLMs)

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

Machine Learning (ML)

Large Language Models (LLMs)

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

Machine Learning (ML)

- Subset of AI using algorithms that improve through experience

Large Language Models (LLMs)

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

Machine Learning (ML)

- Subset of AI using algorithms that improve through experience
- Learns patterns from data without explicit programming

Large Language Models (LLMs)

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

Machine Learning (ML)

- Subset of AI using algorithms that improve through experience
- Learns patterns from data without explicit programming

Large Language Models (LLMs)

- AI systems trained on vast text amounts

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

Machine Learning (ML)

- Subset of AI using algorithms that improve through experience
- Learns patterns from data without explicit programming

Large Language Models (LLMs)

- AI systems trained on vast text amounts
- Understand and generate human language

Artificial Intelligence, ML, and LLMs: Key Definitions

Artificial Intelligence (AI)

- Computer systems performing tasks requiring human intelligence
- Examples: Chess playing, image recognition, language translation

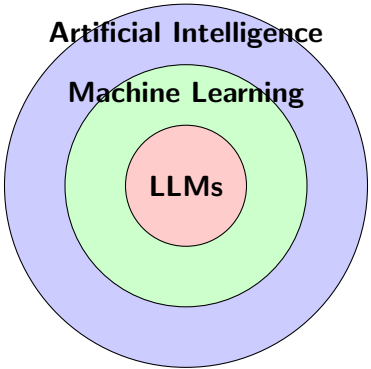
Machine Learning (ML)

- Subset of AI using algorithms that improve through experience
- Learns patterns from data without explicit programming

Large Language Models (LLMs)

- AI systems trained on vast text amounts
- Understand and generate human language
- Examples: GPT, Claude, Qwen, Mistral

AI vs ML vs LLM: The Relationship



LLMs are a specific type of ML,
which is a subset of AI

Model Learning Process: Training Phase

Training Phase

- Model learns patterns from large datasets

Example: LLM learns that "cat" often appears with "dog" in pet-related contexts

Model Learning Process: Training Phase

Training Phase

- Model learns patterns from large datasets
- Adjusts internal parameters to minimize prediction errors

Example: LLM learns that "cat" often appears with "dog" in pet-related contexts

Model Learning Process: Training Phase

Training Phase

- Model learns patterns from large datasets
- Adjusts internal parameters to minimize prediction errors
- Requires massive computational resources

Example: LLM learns that "cat" often appears with "dog" in pet-related contexts

Model Learning Process: Training Phase

Training Phase

- Model learns patterns from large datasets
- Adjusts internal parameters to minimize prediction errors
- Requires massive computational resources
- Can take weeks or months for large models

Example: LLM learns that "cat" often appears with "dog" in pet-related contexts

Model Learning Process: Validation Phase

Validation Phase

- Test model on new, unseen data

Key insight: Good models generalize to new data, not just memorize training examples

Model Learning Process: Validation Phase

Validation Phase

- Test model on new, unseen data
- Measure performance using metrics (accuracy, precision, recall)

Key insight: Good models generalize to new data, not just memorize training examples

Model Learning Process: Validation Phase

Validation Phase

- Test model on new, unseen data
- Measure performance using metrics (accuracy, precision, recall)
- Prevent overfitting (memorizing training data)

Key insight: Good models generalize to new data, not just memorize training examples

Model Learning Process: Validation Phase

Validation Phase

- Test model on new, unseen data
- Measure performance using metrics (accuracy, precision, recall)
- Prevent overfitting (memorizing training data)
- Tune hyperparameters for optimal performance

Key insight: Good models generalize to new data, not just memorize training examples

Model Learning Process: Biases

Common Biases in LLMs

- **Training data bias:** Reflects biases in source material

Real-world impact: Biased models can perpetuate harmful stereotypes

Model Learning Process: Biases

Common Biases in LLMs

- **Training data bias:** Reflects biases in source material
- **Representation bias:** Underrepresents certain groups

Real-world impact: Biased models can perpetuate harmful stereotypes

Model Learning Process: Biases

Common Biases in LLMs

- **Training data bias:** Reflects biases in source material
- **Representation bias:** Underrepresents certain groups
- **Confirmation bias:** Reinforces existing beliefs

Real-world impact: Biased models can perpetuate harmful stereotypes

Model Learning Process: Biases

Common Biases in LLMs

- **Training data bias:** Reflects biases in source material
- **Representation bias:** Underrepresents certain groups
- **Confirmation bias:** Reinforces existing beliefs
- **Temporal bias:** Becomes outdated as language/society changes

Real-world impact: Biased models can perpetuate harmful stereotypes

What is Generative AI?

Definition: AI that creates new content (text, images, code, audio)
Key Features:

- Creates novel outputs, not just classifies

Examples: ChatGPT (text), DALL-E (images), CodeT5 (code)

What is Generative AI?

Definition: AI that creates new content (text, images, code, audio)
Key Features:

- Creates novel outputs, not just classifies
- Learns patterns from training data

Examples: ChatGPT (text), DALL-E (images), CodeT5 (code)

What is Generative AI?

Definition: AI that creates new content (text, images, code, audio)
Key Features:

- Creates novel outputs, not just classifies
- Learns patterns from training data
- Generates human-like content

Examples: ChatGPT (text), DALL-E (images), CodeT5 (code)

Generative AI vs. Other Models

Feature	Generative	Discriminative
Primary Task	Create new content	Classify existing data
Output Type	Novel text/images	Labels/categories
Learning Focus	Data distribution	Decision boundaries
Example	GPT, DALL-E	Spam filters, medical diagnosis

Activity: True or False?

Instructions: Raise your hand if you think the statement is TRUE

Statement 1: Generative AI can only work with text

Statement 2: All generative models are also discriminative

Statement 3: Generative AI creates completely original content never seen in training

Statement 4: LLMs are a type of generative AI

Limitations and Risks: Hallucinations

What are hallucinations?

- LLMs generate false or nonsensical information

Example: Asking for "recent studies on X" might return made-up citations
Risk: Researchers might unknowingly use false information

Limitations and Risks: Hallucinations

What are hallucinations?

- LLMs generate false or nonsensical information
- Appears confident and plausible

Example: Asking for "recent studies on X" might return made-up citations
Risk: Researchers might unknowingly use false information

Limitations and Risks: Hallucinations

What are hallucinations?

- LLMs generate false or nonsensical information
- Appears confident and plausible
- Can include fake citations, non-existent facts

Example: Asking for "recent studies on X" might return made-up citations
Risk: Researchers might unknowingly use false information

Limitations and Risks: Bias

Types of Bias

- **Gender bias:** Assumes doctors are male, nurses are female

Activity: Can you identify bias in this LLM output?

Limitations and Risks: Bias

Types of Bias

- **Gender bias:** Assumes doctors are male, nurses are female
- **Racial bias:** Associates certain groups with negative stereotypes

Activity: Can you identify bias in this LLM output?

Limitations and Risks: Bias

Types of Bias

- **Gender bias:** Assumes doctors are male, nurses are female
- **Racial bias:** Associates certain groups with negative stereotypes
- **Cultural bias:** Primarily represents Western perspectives

Activity: Can you identify bias in this LLM output?

Limitations and Risks: Bias

Types of Bias

- **Gender bias:** Assumes doctors are male, nurses are female
- **Racial bias:** Associates certain groups with negative stereotypes
- **Cultural bias:** Primarily represents Western perspectives
- **Academic bias:** Favors published, English-language sources

Activity: Can you identify bias in this LLM output?

Limitations and Risks: Output Reliability

Reliability Issues

- **Inconsistency:** Same prompt can yield different answers

Best practice: Always verify critical information from multiple sources

Limitations and Risks: Output Reliability

Reliability Issues

- **Inconsistency:** Same prompt can yield different answers
- **Context dependency:** Performance varies by domain

Best practice: Always verify critical information from multiple sources

Limitations and Risks: Output Reliability

Reliability Issues

- **Inconsistency:** Same prompt can yield different answers
- **Context dependency:** Performance varies by domain
- **Temporal drift:** Knowledge becomes outdated

Best practice: Always verify critical information from multiple sources

Limitations and Risks: Output Reliability

Reliability Issues

- **Inconsistency:** Same prompt can yield different answers
- **Context dependency:** Performance varies by domain
- **Temporal drift:** Knowledge becomes outdated
- **Prompt sensitivity:** Small changes affect outputs significantly

Best practice: Always verify critical information from multiple sources

Limitations and Risks: Ethics

Ethical Concerns

- **Data:** Training data may include personal or copyrighted material

Limitations and Risks: Ethics

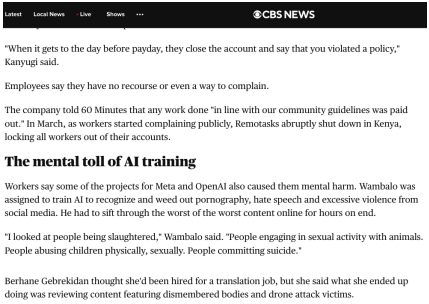
Ethical Concerns

- **Data:** Training data may include personal or copyrighted material
- **Environmental impact:** High energy consumption

Limitations and Risks: Ethics

Ethical Concerns

- **Data:** Training data may include personal or copyrighted material
- **Environmental impact:** High energy consumption
- **Labor practices:** Data annotation often involves poor working conditions



Latest Local News Live Shows ··· CBS NEWS

"When it gets to the day before payday, they close the account and say that you violated a policy," Kanyugi said.

Employees say they have no recourse or even a way to complain.

The company told 60 Minutes that any work done "in line with our community guidelines was paid out." In March, as workers started complaining publicly, Remotasks abruptly shut down in Kenya, locking all workers out of their accounts.

The mental toll of AI training

Workers say some of the projects for Meta and OpenAI also caused them mental harm. Wambalo was assigned to train AI to recognize and weed out pornography, hate speech and excessive violence from social media. He had to sift through the worst of the worst content online for hours on end.

"I looked at people being slaughtered," Wambalo said. "People engaging in sexual activity with animals. People abusing children physically, sexually. People committing suicide."

Berhane Gebrekidan thought she'd been hired for a translation job, but she said what she ended up doing was reviewing content featuring dismembered bodies and drone attack victims.

`https://www.cbsnews.com/news/ai-work-kenya-exploitation-60-minutes` (Nov 2024)

Limitations and Risks: Ethics

Ethical Concerns

- **Data:** Training data may include personal or copyrighted material
- **Environmental impact:** High energy consumption
- **Labor practices:** Data annotation often involves poor working conditions
- **Academic integrity:** attribution of AI assistance / uploading restricted data

Limitations and Risks: Ethics

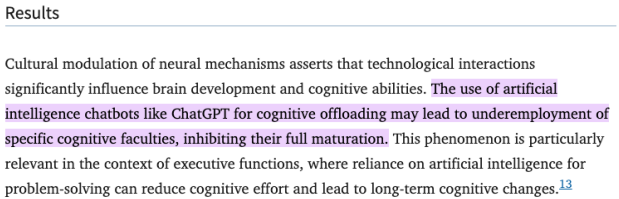
Ethical Concerns

- **Data:** Training data may include personal or copyrighted material
- **Environmental impact:** High energy consumption
- **Labor practices:** Data annotation often involves poor working conditions
- **Academic integrity:** attribution of AI assistance / uploading restricted data
- **Open source:** Often little documentation on data, training, architecture
  "Given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar"
  OpenAI: Achiam, J. et al (2024). *GPT-4 Technical Report*.
  `https://arxiv.org/abs/2303.08774`

Limitations and Risks: Ethics

Ethical Concerns

- **Data:** Training data may include personal or copyrighted material
- **Environmental impact:** High energy consumption
- **Labor practices:** Data annotation often involves poor working conditions
- **Academic integrity:** attribution of AI assistance / uploading restricted data
- **Open source:** Often little documentation on data, training, architecture
- **Effect on cognition:**



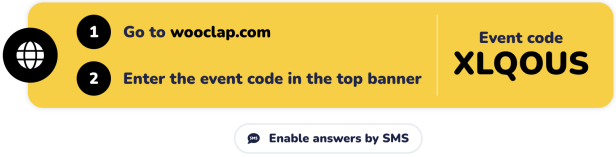
Results

Cultural modulation of neural mechanisms asserts that technological interactions significantly influence brain development and cognitive abilities. The use of artificial intelligence chatbots like ChatGPT for cognitive offloading may lead to underemployment of specific cognitive faculties, inhibiting their full maturation. This phenomenon is particularly relevant in the context of executive functions, where reliance on artificial intelligence for problem-solving can reduce cognitive effort and lead to long-term cognitive changes.[13]

Dubey et al. Redefining Cognitive Domains in the Era of ChatGPT: A Comprehensive Analysis of Artificial Intelligence's Influence and Future Implications. Med Res Arch. 2024 Jun;12(6):5383

Free vs. Paid Systems

Join this Wooclap event



Free vs. Paid Systems

Feature	Free	Paid
Cost	€0	€10-€100+ monthly
Usage Limits	Yes	Higher/none
Model Quality	Basic	Advanced
Response Speed	Slower	Faster
Support	Community	Professional

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting

Choose PAID when:

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use

Choose PAID when:

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints

Choose PAID when:

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints
- Simple tasks (summarization, basic Q&A)

Choose PAID when:

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints
- Simple tasks (summarization, basic Q&A)

Choose PAID when:

- Professional/research work

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints
- Simple tasks (summarization, basic Q&A)

Choose PAID when:

- Professional/research work
- High-volume usage

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints
- Simple tasks (summarization, basic Q&A)

Choose PAID when:

- Professional/research work
- High-volume usage
- Need for reliability

Free vs. Paid: When to Choose What?

Choose FREE when:

- Learning or experimenting
- Occasional, non-critical use
- Budget constraints
- Simple tasks (summarization, basic Q&A)

Choose PAID when:

- Professional/research work
- High-volume usage
- Need for reliability
- Advanced features required

LLM Use in Clinical Psychology Research

Appropriate & Inappropriate Uses

- Think-Pair-Share

LLM Use in Clinical Psychology Research: Appropriate Applications

Appropriate Uses

- **Literature review:** Summarizing papers (with verification)

Key: Always maintain human oversight and verification

LLM Use in Clinical Psychology Research: Appropriate Applications

Appropriate Uses

- **Literature review:** Summarizing papers (with verification)
- **Writing assistance:** Drafting, editing, proofreading

Key: Always maintain human oversight and verification

LLM Use in Clinical Psychology Research: Appropriate Applications

Appropriate Uses

- **Literature review:** Summarizing papers (with verification)
- **Writing assistance:** Drafting, editing, proofreading
- **Code generation:** Writing analysis scripts

Key: Always maintain human oversight and verification

LLM Use in Clinical Psychology Research: Appropriate Applications

Appropriate Uses

- **Literature review:** Summarizing papers (with verification)
- **Writing assistance:** Drafting, editing, proofreading
- **Code generation:** Writing analysis scripts
- **Brainstorming:** Generating research ideas

Key: Always maintain human oversight and verification

LLM Use in Clinical Psychology Research: Appropriate Applications

Appropriate Uses

- **Literature review:** Summarizing papers (with verification)
- **Writing assistance:** Drafting, editing, proofreading
- **Code generation:** Writing analysis scripts
- **Brainstorming:** Generating research ideas
- **Translation:** Working with multilingual sources

Key: Always maintain human oversight and verification

LLM Use in Clinical Psychology Research: Inappropriate Applications

Avoid Using LLMs For:

- **Statistical analysis:** LLMs can't perform proper statistical tests

LLM Use in Clinical Psychology Research: Inappropriate Applications

Avoid Using LLMs For:

- **Statistical analysis:** LLMs can't perform proper statistical tests
- **Fact-checking:** They may hallucinate information

LLM Use in Clinical Psychology Research: Inappropriate Applications

Avoid Using LLMs For:

- **Statistical analysis:** LLMs can't perform proper statistical tests
- **Fact-checking:** They may hallucinate information
- **Critical decisions:** Participant safety, ethical approvals

LLM Use in Clinical Psychology Research: Inappropriate Applications

Avoid Using LLMs For:

- **Statistical analysis:** LLMs can't perform proper statistical tests
- **Fact-checking:** They may hallucinate information
- **Critical decisions:** Participant safety, ethical approvals
- **Original thinking:** Generating truly novel theories

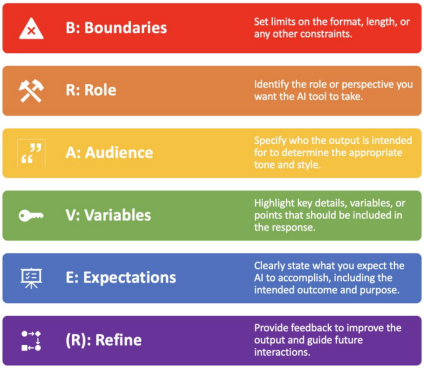
LLM Use in Clinical Psychology Research: Inappropriate Applications

Avoid Using LLMs For:

- **Statistical analysis:** LLMs can't perform proper statistical tests
- **Fact-checking:** They may hallucinate information
- **Critical decisions:** Participant safety, ethical approvals
- **Original thinking:** Generating truly novel theories
- **Replacing expertise:** They supplement but don't replace researcher knowledge

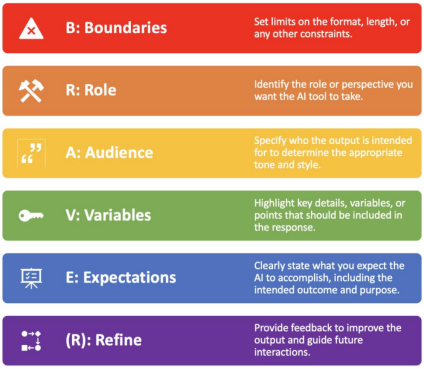
Effective Prompting: BRAVE(R) Framework

Effective Prompting: BRAVE(R) Framework



Prompt formulation by Christine Fox, in collaboration with ChatGPT (May 2024).

Effective Prompting: BRAVE(R) Framework



Prompt formulation by Christine Fox, in collaboration with ChatGPT (May 2024).

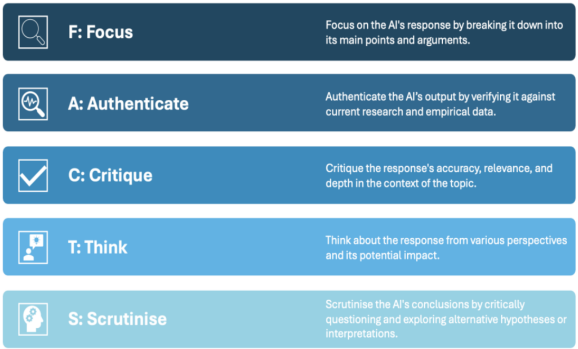
Poor prompt: "Summarize this paper"

Better prompt: "Act as a psychology research assistant [R]. Summarize this paper in 3 bullet points [B] focusing on methodology and findings relevant to cognitive behavioral therapy [V]. Use academic language [A] and APA style [E]."

Key insight: Specific, contextual prompts yield better results

Evaluating and Critically Assessing GenAI Responses

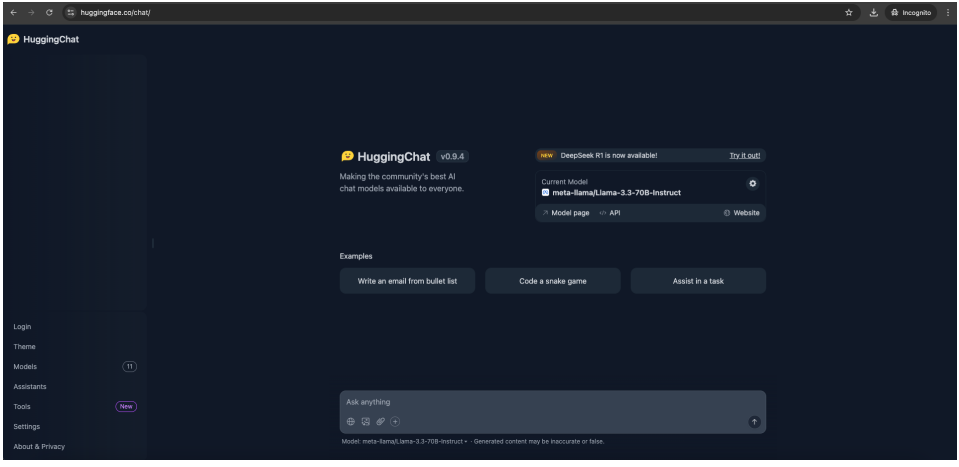
Prompt formulation by Christine Fox, in collaboration with ChatGPT (May 2024).



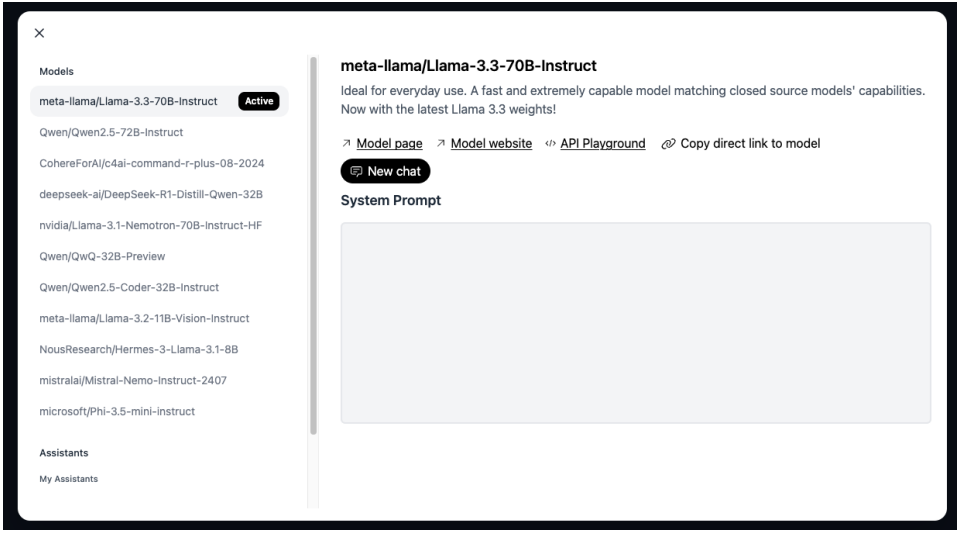
More info:

https://teachersguidegsls.nl/genai/generative-ai-guidelines/

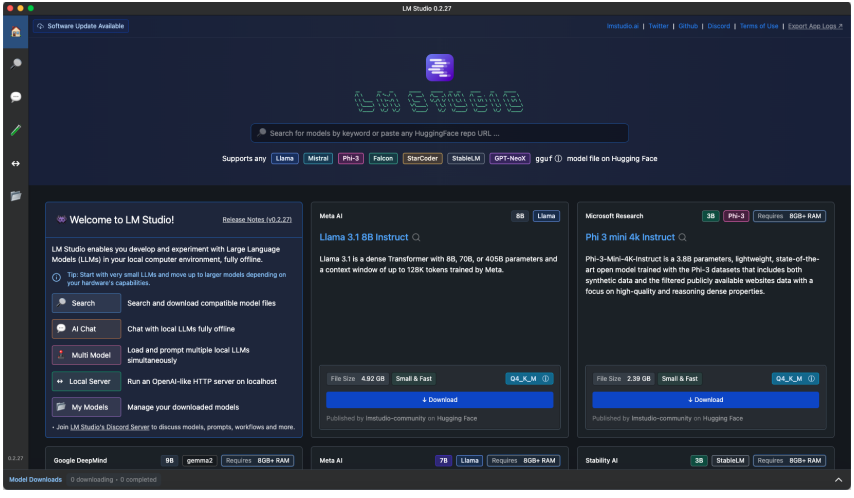
Ways to prompt



Switch models

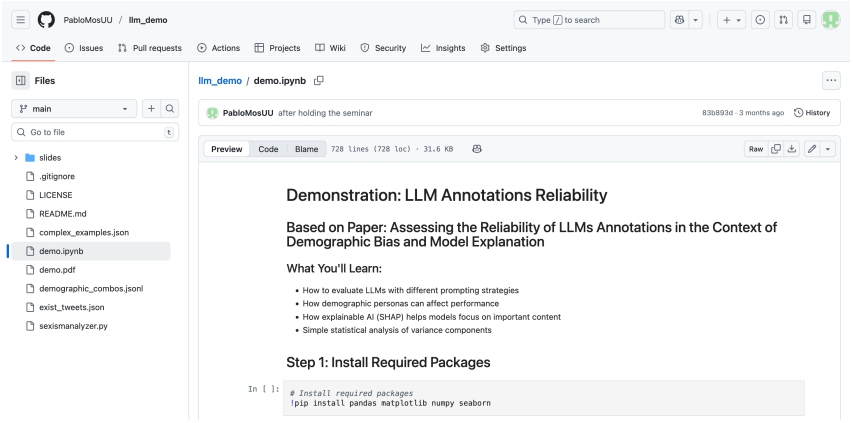


Using an LLM locally (no server)



Versatile platform

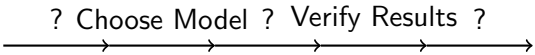
- OpenRouter.ai



https://github.com/PabloMosUU/llm_demo

Activity: Complete the Workflow Diagram

Instructions: Fill in the missing parts of this LLM research workflow



Discussion: What steps are missing? How would you complete this?

Key Takeaways: LLM Workflow

Final Thoughts

Remember:

- LLMs are powerful tools, not replacements for human expertise

Happy researching with AI!

Final Thoughts

Remember:

- LLMs are powerful tools, not replacements for human expertise
- Always prioritize ethics, privacy, and transparency

Happy researching with AI!

Final Thoughts

Remember:

- LLMs are powerful tools, not replacements for human expertise
- Always prioritize ethics, privacy, and transparency
- Verify outputs and maintain critical thinking

Happy researching with AI!