Data concerns and considerations

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Vidhi Chugh
AI strategist and ethicist

Data considerations

- Data volume and compute power
- Data quality
- Labeling
- Bias
- Privacy

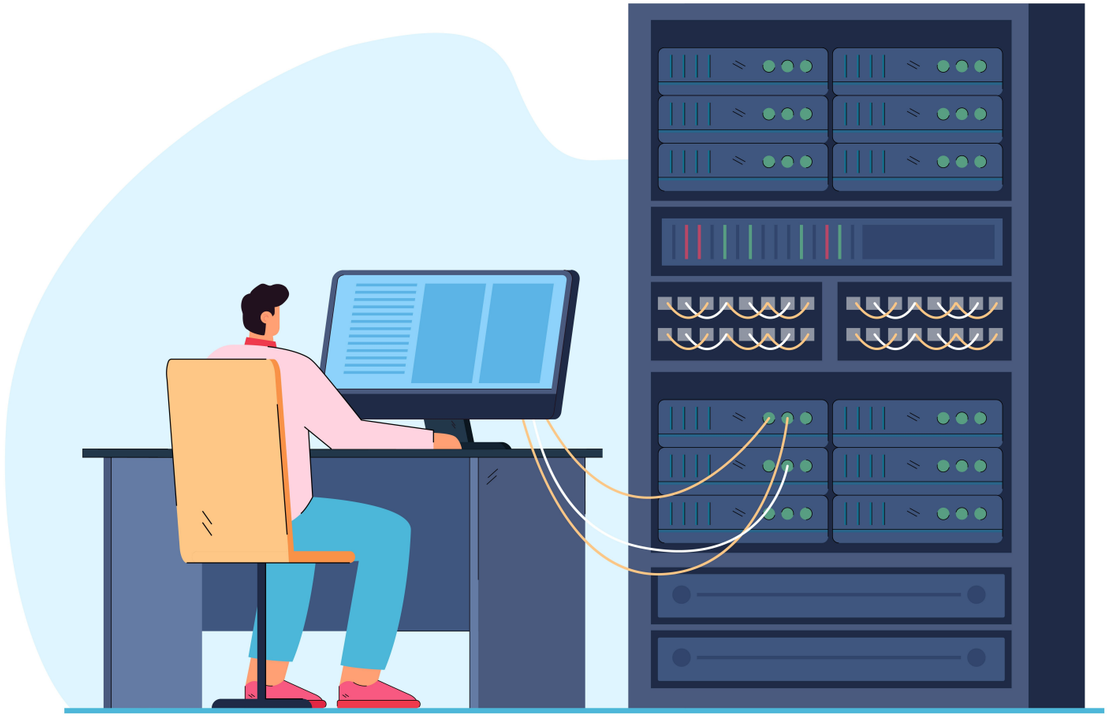
Data volume and compute power

- LLMs need a lot of data
  - Similar to a child learning to talk
  - 570 GB, ~1.3 million books

[1] Freepik

Data volume and compute power

- LLMs need a lot of data
  - Similar to a child learning to talk
  - 570 GB, ~1.3 million books
- Extensive computing power; think of the energy consumption
- Can cost millions of dollars!

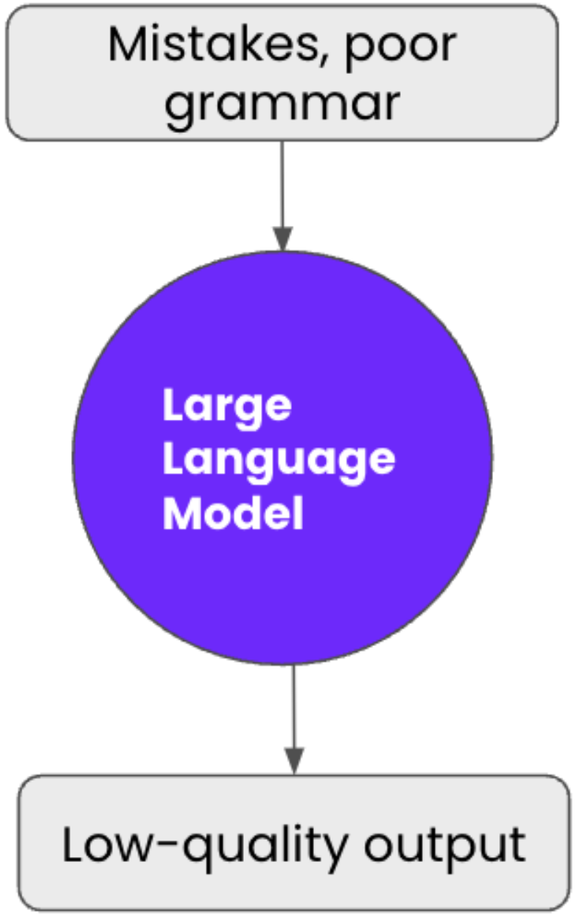


Data quality

- Quality data is essential

- Accurate data = better learning = improved response quality = increased trust

- A child learning to talk
  - Gibberish-in -> gibberish-out



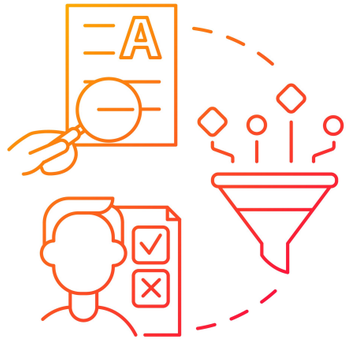
Labeled data

- Correct data label: accurate learning, generalize patterns, accurate responses
- Labor-intensive: assigning correct label to each article

- Incorrect labels impact model performance
- Address errors: identify -> analyze -> iterate

Data bias

- Influenced by societal stereotypes
- Lack of diversity in training data
- Discrimination and unfair outcomes



BIAS

- Spot and deal with the biased data
  - Evaluate data imbalances
  - Promote diversity
  - Bias mitigation techniques: more diverse examples

- Example:
  - "The nurse said that..." -> "she" or "her"

Data privacy

- Compliance with data protection and privacy regulations
- Sensitive or personally identifiable information (PII)
- Privacy is a concern
  - Training on data without permission can lead to a breach
  - Legal, financial and reputational harm
- Get permission

Let's practice!

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Ethical and environmental concerns

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Vidhi Chugh
AI strategist and ethicist

Ethical concerns

- Transparency risk
-
-

Ethical concerns

- Transparency risk
- Accountability risk -

Ethical concerns

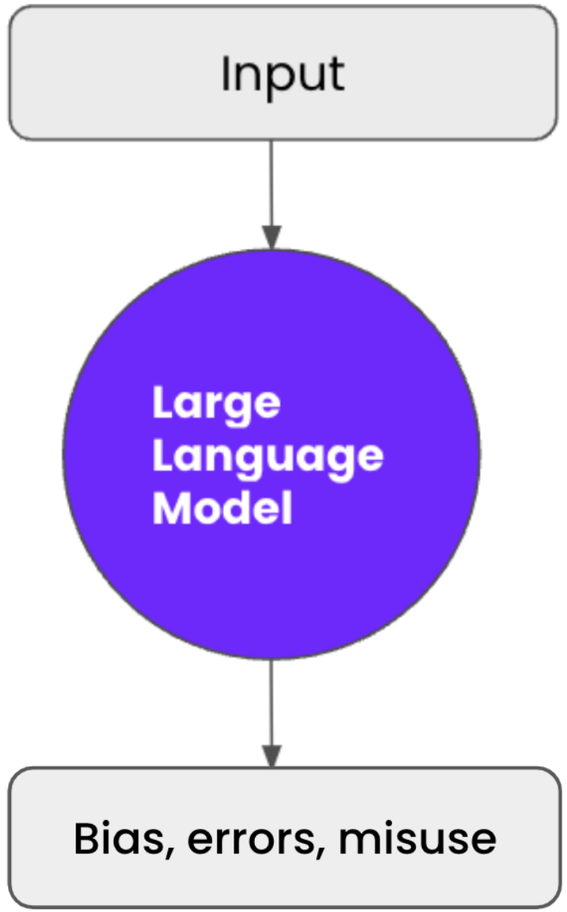
- Transparency risk
- Accountability risk
- Information hazards

Transparency risk

- Challenging to understand the output

- Difficult to identify issues
  - Bias
  - Errors
  - Misuse
- Black box

- Example: reasoning behind predicting disease outcomes



Accountability risk

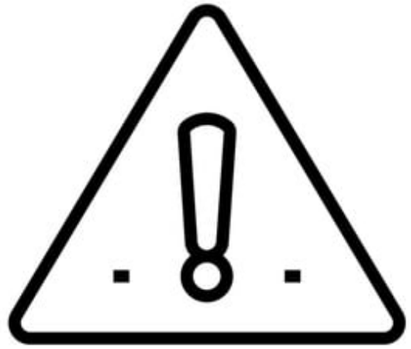
- Responsibility of LLMs' actions

- Who is responsible?
  - Incorrect and harmful advice
  - Model developer or the company?

- Game without rules
  - No transparency
  - No accountability

[1] Freepik

Information hazards



Disseminating harmful information

- Harmful content generation
- Misinformation spread
- Malicious use
- Toxicity

Information hazards

Harmful content generation

- Harmful, offensive, or inappropriate
- Prompt or biased training data
- Example:
  - Bullying vs. friendly school environment
  - Distressing and harmful

Misinformation spread

- Generate text on any topic
- But, no verification!
- Example:
  - "What's a good diet for losing weight?"
  - Unsubstantiated diet plan

Information hazards

Malicious use

- Bad actors exploiting LLMs
- Generate deceptive content
- Example:
  - Fabricated news
  - Manipulating public and causing unrest

Toxicity

- Inappropriate content
- Training or through manipulated prompts
- Example:
  - Insensitive response
  - Stereotype

Environmental concerns

- Ecological footprint of LLMs
- Substantial energy resources to train
- Impact through carbon emissions

[1] Freepik

Cooling requires electricity too!

- Produce considerable heat that needs cooling

- Imagine thousands of laptops overheating
  - Require complex cooling systems
  - Adds to environmental impact

- Balance the cost and benefits
  - Use renewable energy
  - Energy-efficient tech

[1] Freepik

Let's practice!

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Where are LLMs heading?

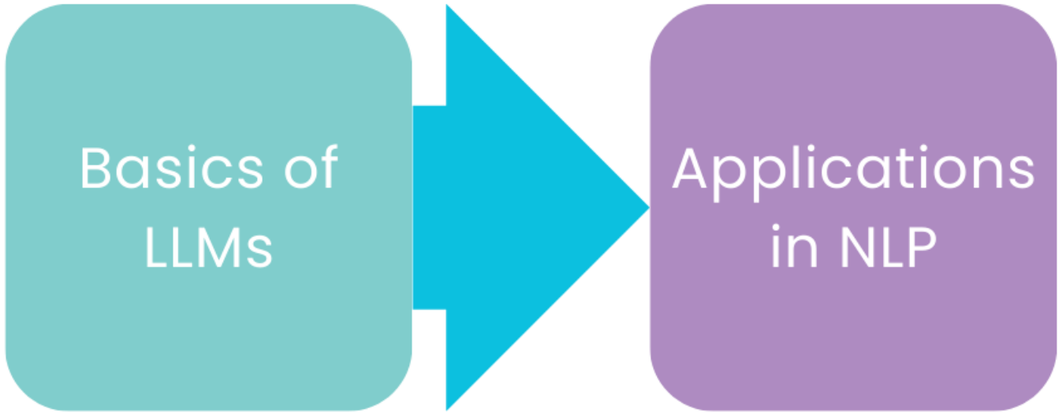
LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Vidhi Chugh
AI strategist and ethicist

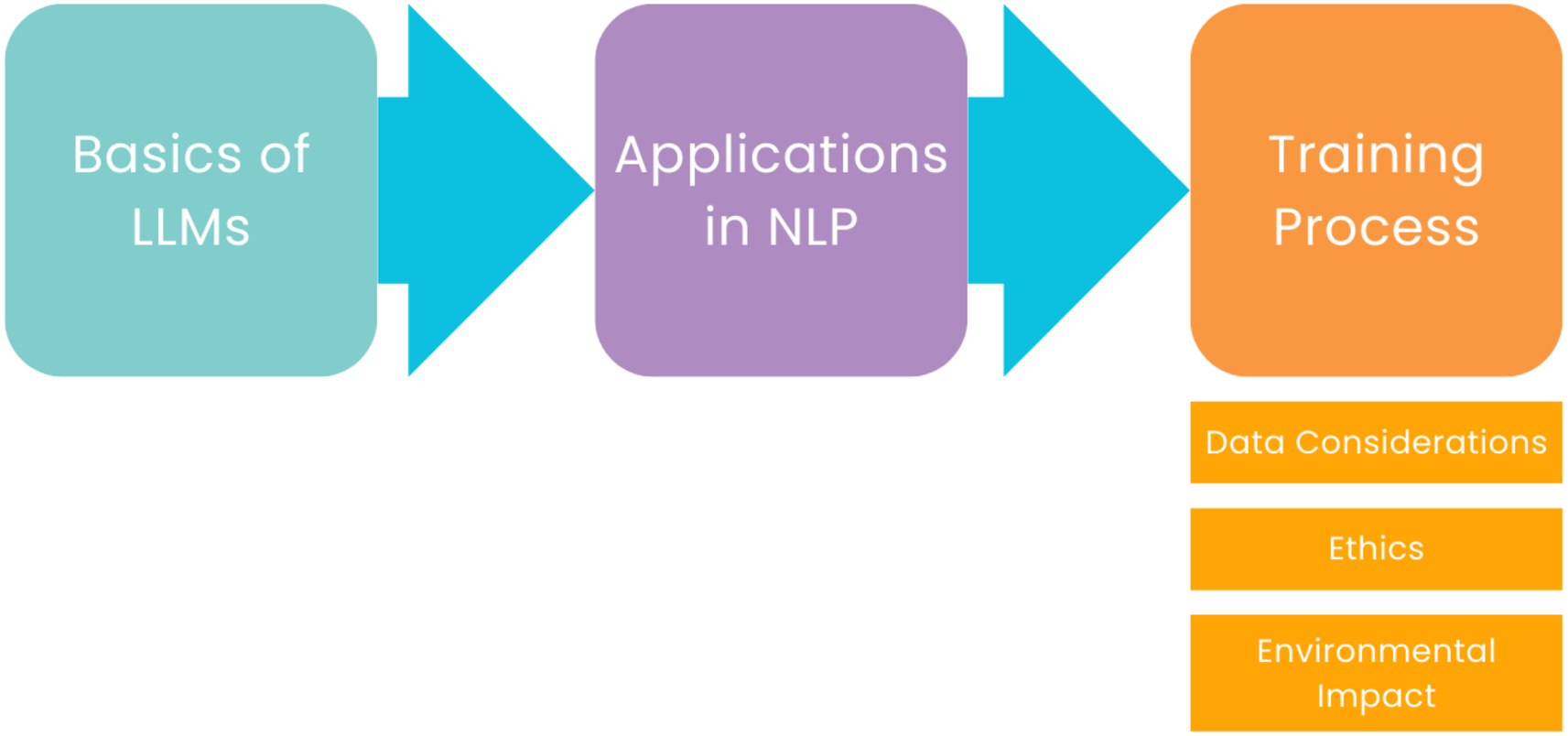
Journey so far

Basics of LLMs

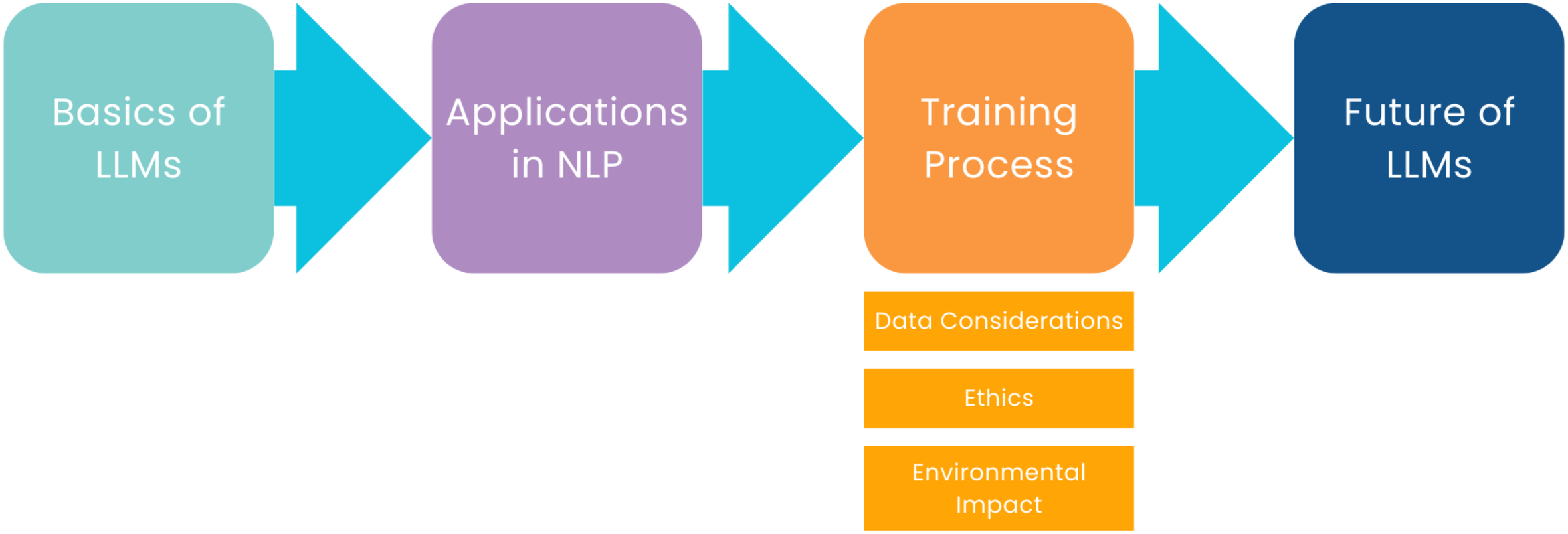
Journey so far



Journey so far



Journey so far



Model explainability

- How do they arrive at their outputs?

- Road-trip planning
  - Why this particular route?
  - Why these specific spots?

- Builds trust and transparency
- Identify and correct the biases or errors

[1] Freepik

Efficiency

- **Computational efficiency**
  - High-quality output with less compute
- **Faster and efficient**
  - Model compression
  - Optimization
- **Benefits:** better storage, lower energy use
- **Accessibility and sustainability**
  - Promotes green AI
  - Reduces operating costs

[1] Freepik

Unsupervised bias handling

- Biased data -> discrimination
- **Unsupervised bias handling**
  - Bias detection and mitigation techniques, automatically
  - No need of explicit human-labeled data
  - Identifies and reduces by analyzing patterns
- **Challenge**
  - Subtle, difficult to detect
  - Might introduce new biases

Enhanced creativity

- Creativity in text-based and visual art forms
- **Artistic content:** learned patterns, not emotional understanding
- Lack human-like comprehension of art or emotions
- Demonstrate human-like emotional behavior
- **Future:** emotion inference

[1] https://arxiv.org/pdf/2302.09582.pdf

Let's practice!

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

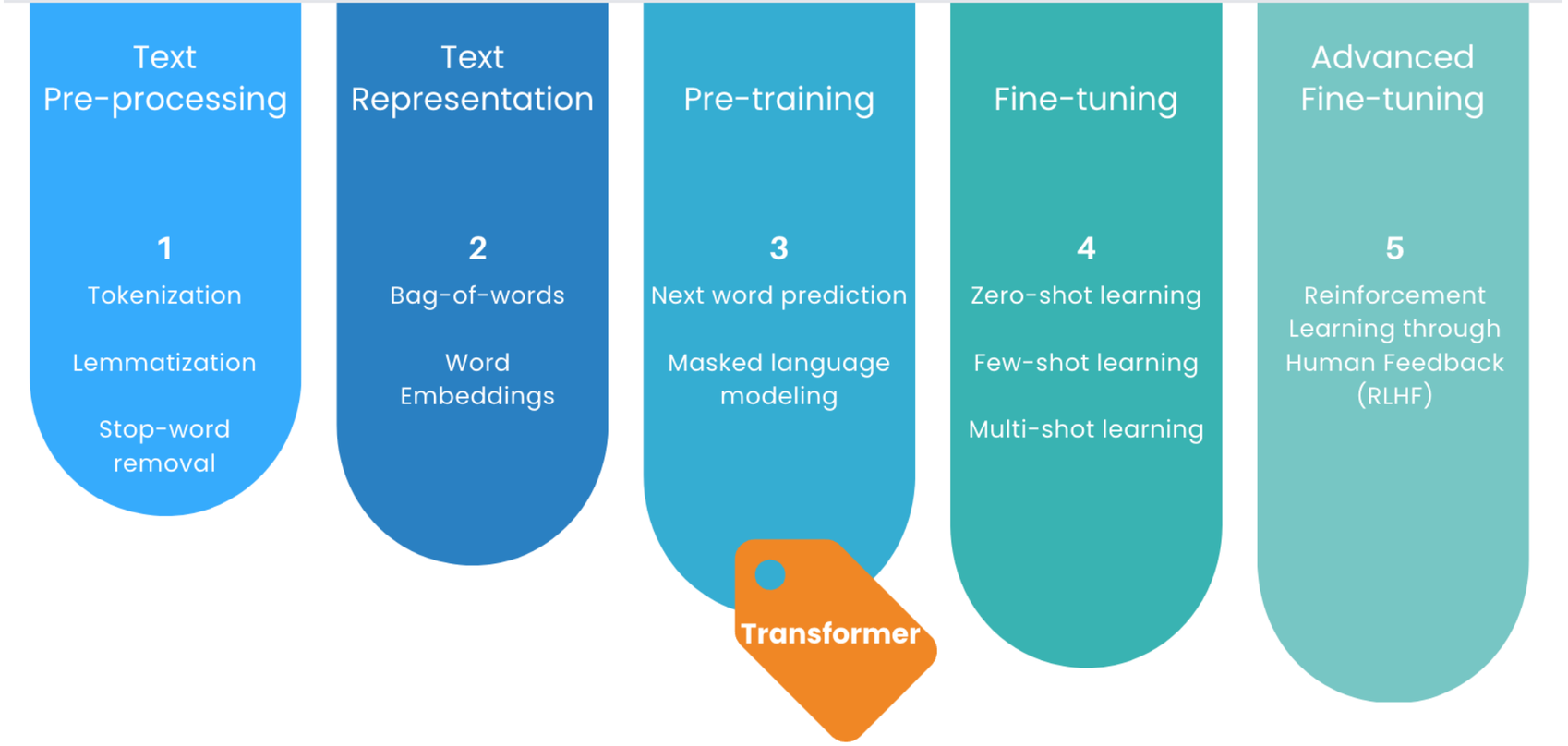
Time to wrap-up

LARGE LANGUAGE MODELS (LLMS) CONCEPTS

Vidhi Chugh
AI strategist and ethicist

How far we have come!

LLMs transforming interaction with technology



How far we have come!

- Substantial data requirements
- Challenges and risks - privacy, ethics, and environmental implications
- Future research and development

There is more to it

- Entire teams devoted to understanding LLMs
- Exciting times ahead
- Stay updated with the latest developments
  - **More on data ethics**
  - **Introduction to ChatGPT**

Congratulations!

LARGE LANGUAGE MODELS (LLMS) CONCEPTS