

# Workshop Outline

Introduction (15 mins)	<ul style="list-style-type: none"><li>• Meet the instructor.</li><li>• Create an account at <a href="https://courses.nvidia.com/join">courses.nvidia.com/join</a></li></ul>
From Deep Learning to Large Language Models (75 mins)	<ul style="list-style-type: none"><li>• Learn how large language models are structured and how to use them:</li><li>• Review deep learning- and class-based reasoning, and see how language modeling falls out of it.</li><li>• Discuss transformer architectures, interfaces, and intuitions, as well as how they scale up and alter to make state-of-the-art LLM solutions.</li></ul>
Break (15 mins)	
Specialized Encoder Models (45 mins)	<ul style="list-style-type: none"><li>• Learn how to look at the different task specifications:</li><li>• Explore cutting-edge HuggingFace encoder models.</li><li>• Use already-tuned models for interesting tasks such as token classification, sequence classification, range prediction, and zero-shot classification.</li></ul>
Break (60 mins)	
Encoder-Decoder Models for Seq2Seq (75 mins)	<ul style="list-style-type: none"><li>• Learn about forecasting LLMs for predicting unbounded sequences:</li><li>• Introduce a decoder component for autoregressive text generation.</li><li>• Discuss cross-attention for sequence-as-context formulations.</li><li>• Discuss general approaches for multi-task, zero-shot reasoning.</li><li>• Introduce multimodal formulation for sequences, and explore some examples.</li></ul>

<p>Decoder Models for Text Generation (45 mins)</p>	<ul style="list-style-type: none"><li>• Learn about decoder-only GPT-style models and how they can be specified and used:</li><li>• Explore when decoder-only is good, and talk about issues with the formation.</li><li>• Discuss model size, special deployment techniques, and considerations.</li><li>• Pull in some large text-generation models, and see how they work.</li></ul>
<p>Break (15 mins)</p>	
<p>Stateful LLMs (60 mins)</p>	<ul style="list-style-type: none"><li>• Learn how to elevate language models above stochastic parrots via context injection:</li><li>• Show off modern LLM composition techniques for history and state management.</li><li>• Discuss retrieval-augmented generation (RAG) for external environment access.</li></ul>
<p>Assessment and Q&amp;A (60 mins)</p>	<ul style="list-style-type: none"><li>• Review key learnings.</li><li>• Take a code-based assessment to earn a certificate.</li></ul>