**Assignment 3.1 - Research & Analysis**

**Understanding Attention Mechanisms in Sequential Models**

**Context**

In Week 2, we introduced the foundations of time series forecasting, dataset exploration, and lookback theory, alongside an early preview of advanced model architectures (LSTMs and GRUs). In Week 3, an Attention Mechanism is introduced into the project. Your task is to investigate attention mechanisms. What they are, how they work, and how they apply to sequential data.

**Project-specific clarification:** In this Nvidia Stock Market Prediction project, we do not implement the classical Bahdanau attention in its original form. Instead, we employ a streamlined “custom simple attention” mechanism, an additive variant inspired by Bahdanau’s approach, tailored for an encoder-only architecture. This adaptation preserves the core additive scoring principle but simplifies the parameterization and application to better align with our multi-lookback GRU submodels, enabling focused weighting of encoder outputs without the complexity of a full sequence-to-sequence framework.

**Assignment Objective**

Research, explain, and critically analyze attention with direct ties to:

* The mechanics of attention (queries, keys, values; scoring + normalization).
* How attention differs from recurrent models without attention.
* How attention, specifically the custom simple attention (additive‑variant), fits into the Week 3 forecasting pipeline.

**Explicit Requirements**

**1) Length**

* **Two pages** (single‑spaced).
* Figures/diagrams are welcome but must fit within the two pages.

**2) Required Section Headings (use these in your paper)**

1. **Introduction**  
   Define attention in deep learning; briefly note its evolution from sequence‑to‑sequence models to broader use in time series.
2. **Core Concept of Attention**
   * Explain queries, keys, values.
   * Describe scoring and normalization (e.g., similarity → softmax).
   * Clarify how attention weights emphasize salient timesteps/features.
3. **Custom Simple Attention (Additive‑Variant) Used in This Project**
   * Explain the adapted additive formulation at a high level (what inputs it consumes; how it produces weights, where the context vector is used).
   * Clearly state how this differs from canonical Bahdanau (e.g., parameterization, scoring simplifications, where it’s applied in the architecture, training implications).
   * Relate this to our multi‑lookback GRU submodels.
4. **Comparison to Non‑Attention RNNs**  
   Contrast with LSTMs/GRUs that rely only on fixed lookbacks; discuss memory limits, vanishing gradients, and long‑range dependency capture.
5. **Applications to Financial Forecasting**  
   Provide **concrete use cases** showing how attention improves:
   * Feature/step importance over long horizons (e.g., earnings windows, volatility spikes).
6. **Conclusion**  
   Summarize the value of attention for Nvidia Stock Market History prediction pipeline and note next steps (e.g., experimenting with Luong or lightweight self‑attention; ablation with/without sentiment).

**3) Research & Sources**

* Cite at least 3 academic or authoritative technical sources.
* APA citation style only.

**4) Format & Submission**

* **Submit as a Word document.**

Evaluation Criteria (100 pts)

| Criteria | Points |
| --- | --- |
| Completeness of required sections | 20 |
| Technical accuracy of attention mechanics | 20 |
| Correct treatment of our project’s “custom simple attention (additive‑variant)” and how it differs from textbook Bahdanau | 15 |
| Depth of analysis linking attention to sequential data forecasting | 20 |
| Quality of comparison with non‑attention RNNs | 10 |
| Writing clarity, organization, and APA citations | 15 |

Tips (keep it concise to stay within three pages)

* Use one small diagram (if any) to illustrate scoring → weights → context vector.
* When contrasting with Bahdanau, be explicit: what’s simplified/changed in our variant, and why that’s suitable for our multi‑lookback GRU setup.
* For multi‑modal fusion, describe how attention could weight recent high‑impact news versus long‑horizon technical context.