AI Seminar Report

Date: 03/14/2025

Name	Vo Hoang Chuong (우황장)	**Student No.**	2024121193
School	Dongguk University	**Course Classification**	Ph.D.
Speaker Name	임성훈	**Speaker Affiliation**	DIGIST
Title	Pushing the Boundaries of Robot Intelligence: Data Scaling by World Foundation Model with Physical Laws		

■ Seminar Contents

- **A Historical Trajectory:**
  - Big early advancement: CNNs (AlexNet, ImageNet - early 2010s) - visual recognition tasks.
  - Deeper networks: ResNet (2015) – increasing accuracy.
  - Sequential modeling: Transformers (2017) – disrupted NLP.
  - Language understanding: BERT (2018) – contextual understanding.
  - Multimodal models: Vision-Language models: CLIP, ViT (2021) – the vision + language bridge.
  - Large Language Models (LLMs) (2022): Generative text capabilities.
- **Currently Focused: Foundation & Adaptation:**
  - Foundation Models: General purpose, pre-trained on large knowledge bases.
  - Adaptation: Specializing foundation models in specific tasks.
  - Benefit: Can learn from fewer specialized data.
- **Next Wave: Data Scaling:**
  - Main Theme: Smart, iterative data selection, scaling more data.
  - Model size increasing: Linear increasing through the years.
  - **Tesla's study case:**
    - An iterative approach: train, identify weaknesses, augment data, label refinement.
    - Simulation data: adversarial models.
- **Generative Models: Powering Data Scaling:**
  - Concept: Models that generate data, not merely analyze.
  - Key Advantage: Opens up enormous potential for data scaling.
- **Nvidia's COSMOS: World Modeling in Action:**
  - World Models: Generative models that comprehend real-world dynamics (physics, space)
  - Goal: Foresee future states from past/present observations.
- **Technical Highlights:**

- Massive Dataset: Millions of hours of internet videos (HD up to 4K).
- Data Curation: Transforming raw video into a trainable format.
- Tokenization: Unsupervised learning (autoencoders) to create a latent space.
- Diffusion Models: Core technology for generating video sequences.
- Autoregressive Model: Iterations of COSMOS.
- Simulation: Nvidia Omniverse for realistic driving data.
- Rigorous Evaluation: Direct comparison of generated and real-world data.

- **The Importance of World Models:**
  - Unlimited Data: Possessing an infinite amount of training/testing data.
  - Controlled Environments: Exact control over variables in synthesized data.
  - Enhanced Realism: Focus on both visual accuracy and physical accuracy.
- **Future Directions:**
  - Data Scaling as Extrapolation: Breaking free from interpolation with existing data.
  - Generative Models & Physical Laws: Anchoring generation of new data points.
  - Cross-Modal Methods: Leveraging text, images, point clouds, and video.
  - Goal: AI that understands the world's structure and dynamics.

■ What have you learned from this seminar?

- **Deep Learning's Journey:**
  - We've witnessed distinct phases, each expanding AI's fundamental capabilities: from early image recognition to sophisticated language understanding and now, generative power. It's a story of qualitative leaps, not just incremental steps.
- **The Power of Shared Knowledge:**
  - These large, pre-trained models act like knowledgeable generalists. Fine-tuning them for specific tasks is like tapping into a vast reservoir of existing understanding. This accelerates learning and reduces the need for massive, specialized datasets.
- **Data:**
  - It's not just about more data but more intelligent data. We need strategic, iterative approaches to collecting, refining, and labelling information. Think of it as a continuous cycle of learning and improvement, like Tesla's approach.
- **Generative Models:**
  - This is a game-changer. Models that can create data, not just analyze it, open up incredible possibilities. Imagine a virtually unlimited supply of training scenarios, allowing us to explore the edges of what's possible.
- **Exploring the Unknown:**
  - We're using generative models, guided by the laws of physics, to create new data points, effectively expanding the boundaries of what AI can learn. It's like venturing into uncharted territory.