



A Study on the Application of LLM-AI for Korean Econometrics Research

Dong Hwa Kim

DSTSTC, Daejeon City, S. Korea

Abstract: LLM among AI has been increased powerful tool that has the potential to revolutionize research as well as general purpose. LLM (large language models) such as ChatGPT based on LLM can assist not only communication, art translation between art and text but also research such as sciences, engineering, economy trend, and forecasting of GDP growth. It enables economists to forecast the growth of GDP by describing all domain data and graph. LLM can also provide implications of LLM-powered cognitive automation for economic research, in some areas. This paper can explain how to get started and provide preparation on the latest capabilities of LLM (AI) in economics. To research on purpose, the paper of LLM should mention about what kind of language models they use, what is the core approaches, what is the tuned parameter (PEFT), and how they obtain practical data for application. These methodologies of LLM application should be provided through research and simulation of real-world LLM applications across the scientific area, engineering field, healthcare issues, and creative topic. The research of Korea econometrics based on LLM (K-EcoLLM) should be performed through a review of current economic situation and data on purpose because we cannot fully and effectively use by the LLM of general purpose. That is why this paper provide the motivation and research strategy on Korean econometric by LLM.

Keywords: GAI, LLM, ChatGPT, Econometrics, Korean Econometrics.

I. INTRODUCTION

Currently, so many are increasing to introduce into creation such as image generation, drawing, photo generation, and storytelling because of the current modal function of ChatGPT (Chat Generative Pre-Trained Transformer).

The ChatGPT was developed for the service of language train function but it has capability after modal function by supervised learning, unsupervised learning, and reinforcement learning. Therefore, its learning mechanism and model are not simple. However, its impact is wide and big everywhere including in science and research domain. Currently, many are developing several models such as LLM, sLLM, and others to have an initiative on economic properties through customer services. Herein, its application and impact will be wider in everywhere including the art generation and movie areas. They are also trying to develop a simple model and widen services in several areas.

As research domains, LLM (Large language model) enable economists to predict from many materials such as newspapers, data, and others. That is, we need to offer advancements in LLM design, training, adaptation, evaluation, and application for both researchers and practitioners for econometric research. To accomplish this, research should mention the evolution of language models, core approaches, PEFT (Parameter-Efficient Fine-tuning), and practical data for application. The methodology of LLM application should mention the survey of real-world LLM applications across the scientific area, engineering field, healthcare issues, and creative topic and method through a review of current example and benchmarks. For those, inference costs of LLM, market structures, raising economic, labor concerns, and benefits of human assistance by using of LLM. The main trend of LLM is doing the development of multimodal and agentic systems for all topics. For example, CLS (Columbia Law School) strives to admit individuals to the LL.M, which is program who are committed to academic and professional excellence through their prior experience, recommendations, and academic history, demonstrate the qualities and skills necessary for rigorous intellectual engagement. Admission decisions are based not only on your potential but also on your demonstrated motivation and self-discipline.

The detailed description [21-25]:

- Forged your values and achieved your prior goals?
- Chosen to commit your time, energies, and talents?
- Made use of your opportunities?

Additionally, this program has an intention to provide that diversity of all kinds deeply enriches the Columbia student body. Their admissions process welcomes applicants of all professional and life experiences.



For instance, University, Harvard University, UC Berkeley, Cambridge University, Northwestern University, UCLA, University of Illinois, University of Chicago, University of London, University of Toronto, NYU, University of Washington, Yale Law School, Stanford University, etc., as master course, of Ph. D or training course. They provide for topics as;

- Advanced European Union Law
- Energy and Climate Law (Top Rated LLM)
- Global Criminal Law
- Health and Technology Law
- International Commercial Law
- International Human Rights Law
- Public International Law (Top Rated LLM)
- Technology Law and Innovation

It means LLM technologies is not option and absolute to study.

For important economy, we should provide an econometric framework that addresses these complexities and provides practical guidance for empirical research using LLM outputs, focusing on prediction problems and estimation problems. In prediction problems of economic issues, researchers use collected text to predict some economic outcome. For this, we usually extract meaning from text and model the complex structure of language. Because LLMs has already capability from enormous training history, it can possibly serve as the foundation upon which economists can predict problems either by directly prompting an LLM to make a prediction or by using its representations.

Basically, when we use LLMs for prediction problems, researcher build form using an LLM and evaluate on an evaluation dataset for predictions. In estimation problems, researchers estimate relationships between economic concepts expressed in text, social media, and downstream parameters. There are some resources for measuring the economic concept. Researchers should use an LLM to economize on measurement costs. How can they use LLM outputs for valid downstream inference? That is a quite important to use LLM in econometrics. If data or prompts lose a validation data, researchers cannot assess the magnitude or pattern of errors in LLM output. Therefore, they will not evaluate their impact on downstream parameter estimates [26].

To use LLM for econometrics practically, how to phrase the prompt to lead different parameter to estimates in applications of finance and political economy with coefficients varying in magnitude, sign, and significance. Especially, for mentioning economic impacts, which impact will increase more market including image generation of art or drawing, movie, news, u-tube market, and its application. Currently, customers' communication patterns have been changed from multi and u-tube styles. It means that everybody can make video stations and communication hubs to build businesses or hobbies easily. That is, as far as they want, they can make image and video art for their economic properties and interests. The art generation market by model will increase more for economy. In June 2023, McKinsey predicts that generative AI will change higher labour productivity, higher education, and higher. It means the role of LLM is bigger and bigger and will be super generative including art areas and will influence economic growth factors. Therefore, this paper provides material for prediction research and motivation to research econometrics.

II. MOTIVATION AND REVIEW OF PREVIOUS RESEARCH ABOUT KOREAN ECONOMETRICS RESEARCH BY LLM

A. Motivation

Currently, workflow and business of industry and service domain have been increasing after the mid-2000s because of Generative, LLM, and ChatGPT. They also have been trying to explore the role of automation and artificial intelligence (AI). Their research highlight is to how machine learning algorithms is applied in their workflow which can optimize the different tasks such as scheduling, manufacturing, and management. Additionally, the rise of the Internet of Things (IoT) and cloud computing provide new advantages for real-time data collection and predictive analytics, further improving risk management and decision-making capabilities.

In research of science and research domain, research process offers how machine learning models can be used to predict project outcomes, planning, automatic routine design tasks, significantly reducing human error, improving project efficiency, and risk prediction. In physical product area, some plan and risk can be complemented by the introduction of augmented reality (AR) and virtual reality (VR).



The milestone of generative AI, large language models (LLMs), and OpenAI's ChatGPT is developed and many areas are introducing so far.

These tools brought advanced natural language processing capabilities to many industries including the specified domain and they enable the automated report generation, multi-lingual communication, and knowledge management. Research has explored how generative AI could address inefficiencies in professional workflows and improve collaboration across diverse teams.

In case of ChatGPT in engineering domain, it has been getting attention because of its ability to generate actionable insights from complex datasets, making it a valuable tool for decision-making in data-intensive projects.

The integration of generative AI into education and training began after 2021 [37]. Several studies show the application of ChatGPT as an educational resource for engineering that foster personalized learning and enhanced critical thinking. Researchers also began integrating generative AI with AR and VR platforms to create immersive training environments. Research should also explore the ethical implications and cybersecurity risks associated with AI adoption, proposing frame-works for responsible implementation.

Additionally, advanced research should present the results through integrating ChatGPT with research project, as well as real-time risk assessments and predictive maintenance [31,34].

Generative AI technologies are being used in sophisticated applications, including automated maintenance tasks, multilingual information retrieval, and dynamic safety assessments [32,41,42].

In econometrics research domain, LLMs perform reasonably well for general purpose. However, the econometric models for LLM should be employed trained using data from the forecast period for a notable achievement, even without information leakage. They show that a fair comparison of the two approaches using an in-sample evaluation of the econometric models yields even more favourable results [39]. They suggest that ready-made LLMs are not yet ready to replace traditional models for operational forecasting and they should provide LLM that can have the valuable additional insights during periods of structural change or heightened uncertainty.

They also find that the language of the prompt has a clear impact and the effects of prompt design (justification versus no justification; direct versus narrative) are less conclusive. (English prompts yield better results: Author tried in French). Nevertheless, the narrative prompt effectively forces the platform to generate nowcasts for certain models. They suggest the work across different experimental settings, including model versions. The analysis is limited to the use of LLMs with prompt queries and no external data.

They are interested in and focus on exploring how LLM performance improves when they enrich the inputs such as Banque de France surveys, newspaper content or time series data. They show Using more advanced prompting techniques or fine-tuning LLMs on current data could further improve their predictive capabilities. Combining LLMs and econometric models could enhance their respective strengths and improve overall forecast accuracy.

Finally, they are interested in cooperating with the United States for robustness since LLMs are trained more intensively on English-language sources. This is why we study the LLM focused on purpose.

B. Overview of AI and ChatGPT

Currently several researchers approach to use AI and GAI (Generative Artificial Intelligence) for economic analysis [2, 3, 14, 15]. Especially, Since LLM (Large Language Model) have the possibility, many try to apply and develop innovatory research in economics or market analysis. The first version of ChatGPT was released by OpenAI in November 2022 and updated in April 2023 [1]. The several event during history of AI and ChatGPT show in Fig. 1-4.

So far, there are many developments for Chat GPT area or similar technology after ChatGPT in 2022. Especially, suddenly advanced functions of its related technology. As LLM, the year of 2025 is the third years in GPT series after 2023, 2024 (Test version in Dec. 2022). By training LLMs against automatically verifiable rewards across a number of environments, the LLMs spontaneously develop strategies that look like reasoning to humans. It means that it can learn to break down problem solving into intermediate calculations and it can learn a number of problem-solving strategies for going back and forth to figure things out. Running RLVR turned out to offer high capability/\$, which gobbled up the

compute that was originally intended for pretraining. That is, there are many progresses and the capability for LLM in 2025.

Of course, many AI researchers released at least one reasoning model in 2025 or hybrids that could be run in reasoning or non-reasoning modes and we can find that more complex research questions can often be answered by GPT-5 Thinking in ChatGPT. Reasoning functions are also increased at producing error and debugging code.

Till 2024, a very few people agent AI but from 2025 everyone was talking about agents and it appeared to be developing. And then so many define agents as LLM systems that can perform useful work via tools. For development researchers, many labs put out the CLI (Command Line Interface) such as Claude Code, Codex CLI, Gemini CLI, Qwen Code, and Mistral Vibe. Their comparison is Fig. 5 [6, 18].

For image learning, OpenAI released an API version of the model of gpt-image-1 and soon opened by a cheaper gpt-image-1-mini in October. They opened an improved gpt-image-1.5 on December 16th, 2024 [37].

The Google image opened in the name of their Nano Banana models for availability via Gemini. They opened their own Gemini 2.0, Gemini 2.5 and then Gemini 3.0 in 2025, which each model supports audio/video/image/text input of 1,000,000+ tokens, priced competitively and proving more capable than the last.

So far, we reviewed the function of ChatGPT and its family model to understand on how we have to use and apply or approach [19].

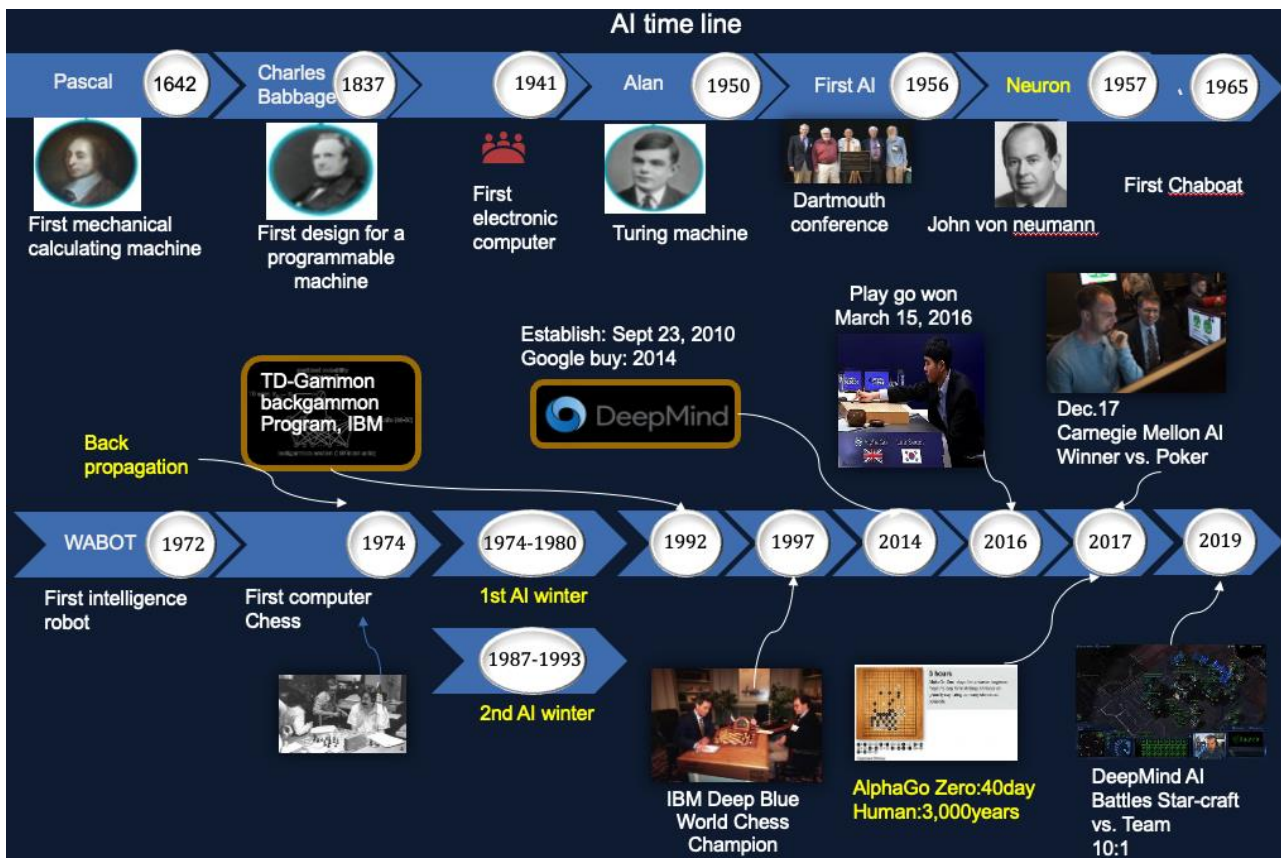


Fig. 1 The brief history of AI to understand ChatGPT [9, 28-30]

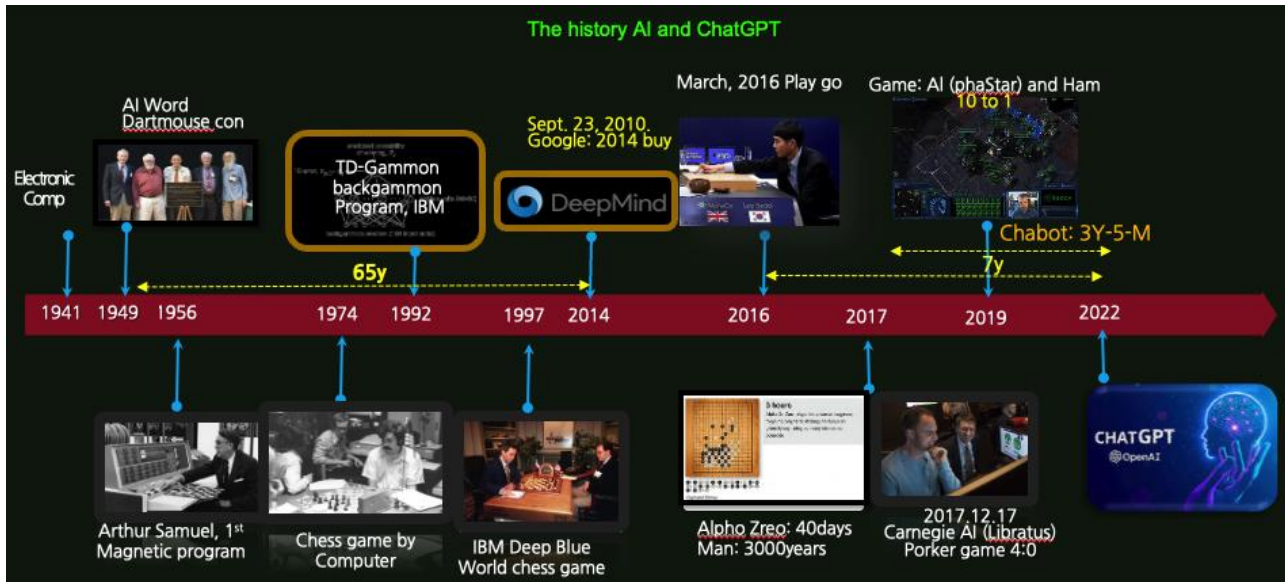


Fig. 2 The brief history of ChatGPT [9, 28-30]

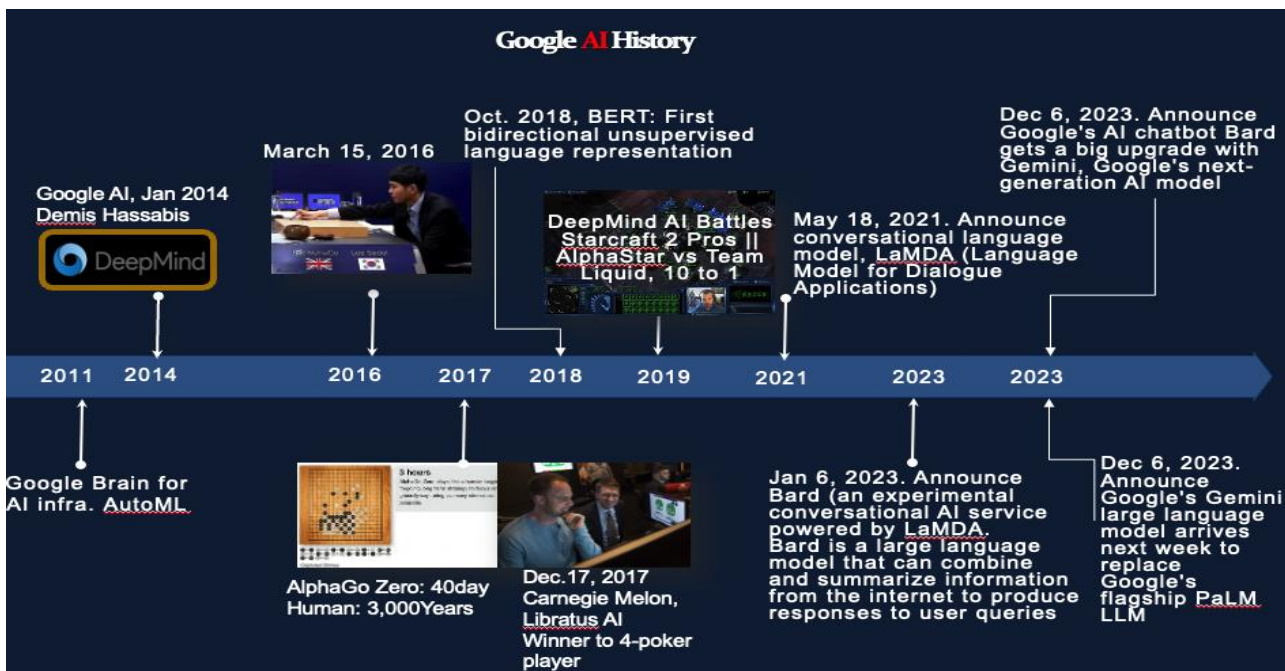


Fig. 3 The brief history of Google AI and ChtGPT [9, 28-30]

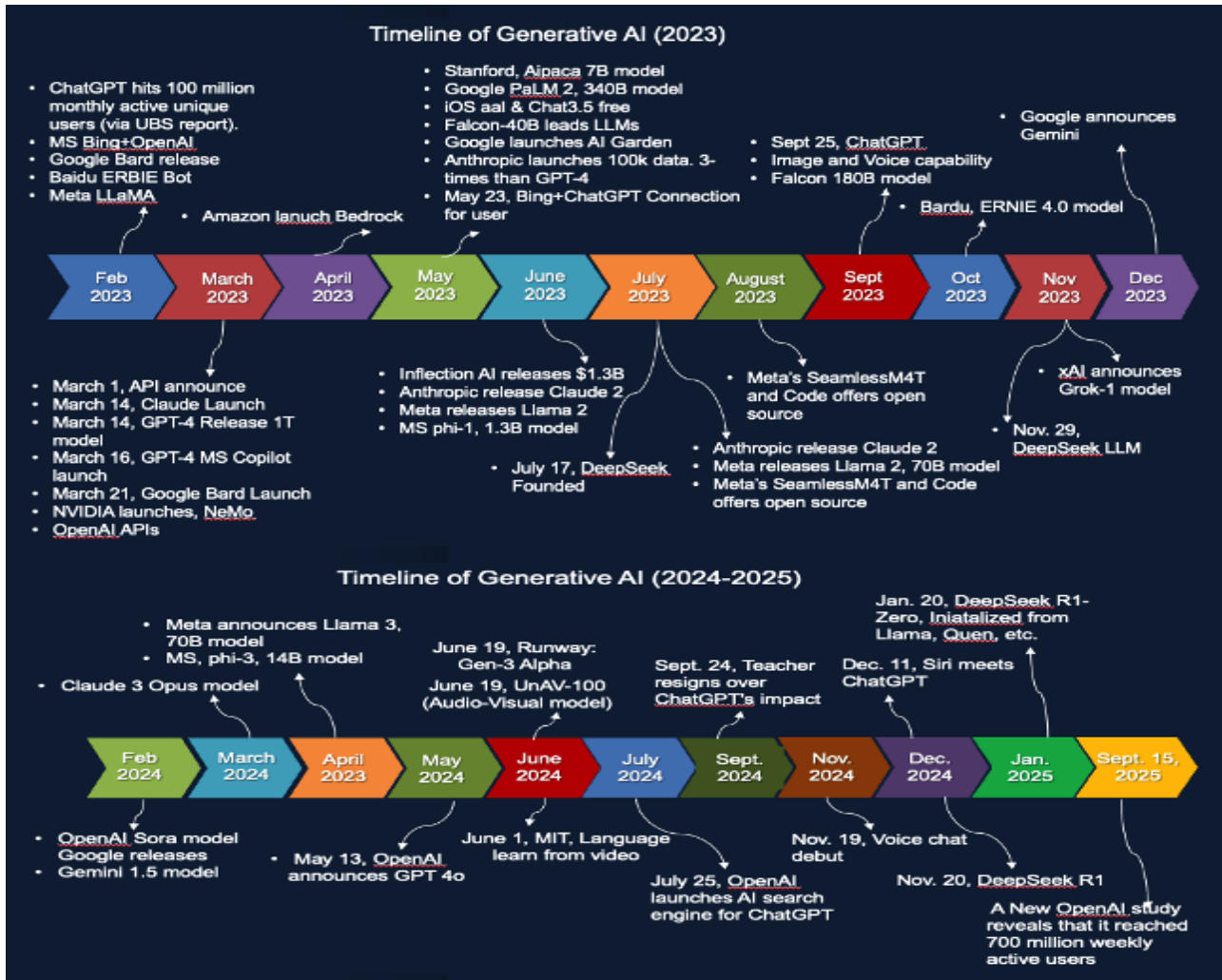


Fig. 4 The brief history of ChatGPT during 2023-2025 event [9, 28-30]

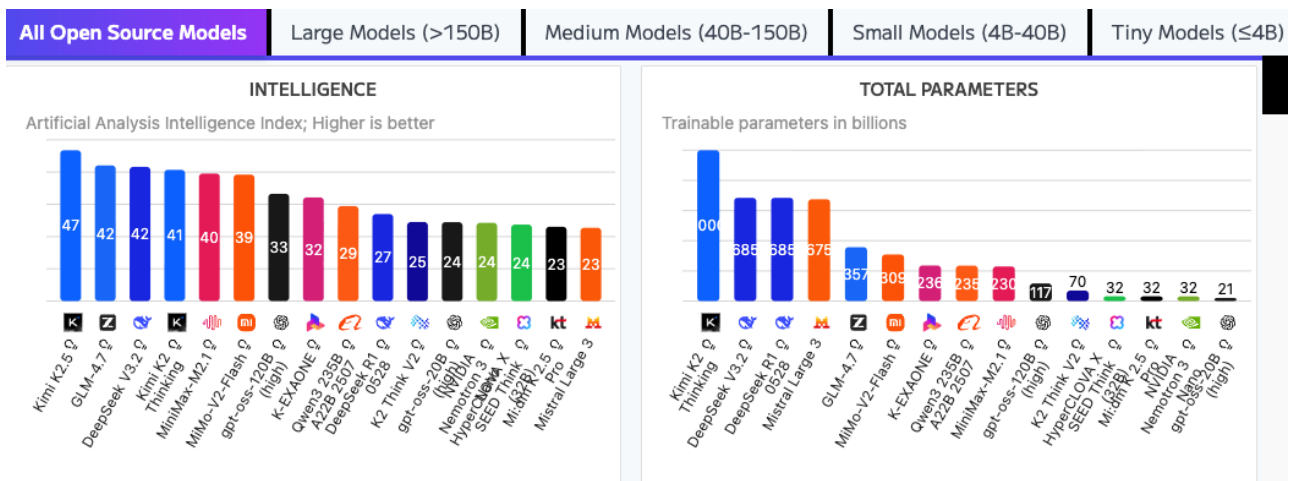


Fig. 5 Comparison of LLM functions and parameter (<https://artificialanalysis.ai/models/open-source>)

TABLE I DEFINITION OF GAI, LLM, AND CHATGPT [30, 37, 38]

Feature	Generative AI	LLM	ChatGPT
Definition	Wide broad field of AI that creates new content (text, image, model) based on input data. (Ex. A car as general transportation)	A model of Gen-AI specialized for producing text/language like human language. (Ex. The engine (specific component))	A conversational model based on LLM for dialogue like chatbot (It is one of service produced using a LLM) (Ex. A Tesla (specific product for specified service)).
Key Feature	Content creation, predictive capability, data synthesis across multiple modal format	Natural language processing, context understanding, Multilingual capability	Interactive text generation, real-time response, task-specific customization.
Output	Text, images, audio, code, video.	Text, code, translations, summaries.	Text, code, and images.
Application Scope	Automated content generation (Manuals, design recommendations), predictive analytics for project planning, simulation of construction site scenarios, real-time data processing for risk assessment, automated compliance reporting, and proactive hazard prevention strategies, Scope included LLM	Multi-lingual communication for global projects, generation of technical reports, analysis of legal documents & contracts, summarising project, real-time question-answering during meetings, integration with BIM for enhanced data insights, and predictive analytics. Subset of Generative AI.	Personalised learning service for education of specified field, AR-based immersive training, automated compliance checks, real-time safety hazard identification, collaborative document editing, support for multi-lingual project teams, generation of daily logs for target purpose (Ex. construction site), real-time safety guidance, and advanced risk management planning, A specific application.
Strength & Limitation	Highly adaptable, versatile across applications, fosters innovation, Requires substantial computational resources, ethical concerns such as data misuse.	Highly accurate language comprehension for supports multilingual collaboration. Can generate inaccurate content (hallucination), requires domain-specific tuning.	User-friendly interface, adaptable for diverse target-related tasks, scalable across applications. Prone to errors in complex queries, limited domain-specific depth without fine-tuning.

C. ChatGPT Application for Economic Research

Table 1 shows the application of LLM. From this table, we can see the power of LLM for science and engineering. That is why this paper mention more research on purpose and target. There are several research materials for economic trend and related topics in the world. Ref. [6] researches market analysis by GAI (Generative Artificial Intelligence). They try to apply its capabilities to data analysis and provide powerful tools for businesses aiming to establish or maintain a strong brand position by GAI [9, 28-30].

This research examines the application of the LLM (Large Language Model) for market research. They aim to evaluate the extent to which GAI can replicate traditional market research. They perform a comprehensive survey on beer consumption in Spain, covering all conversion funnel stages, from Brand Awareness to Purchase.

They study using four prominent LLM such as ChatGPT (OpenAI), Gemini (Google), Claude (Anthropic), and LLaMa (Meta). The results of these LLMs were compared with those of the traditional survey using a collection of statistical methods. Their results show that LLMs are valuable for market research, offering significant insights as reliable proxies. They represent a competitive advantage by making studies of this kind more accessible and cost-effective, benefitting companies of all sizes. However, they mention that LLMs cannot fully replicate traditional methods and present result variability, introducing risks in decision-making because of potential errors in data generation that are complex to estimate without benchmarking.

TABLE III APPLICATION CATEGORY OF LLM

Category	Sector	Description	Ref.
STEM & Research	Scientific area	LLM (Elicit, SciBot), knowledge analysis, planning, scientific writing	40
	Healthcare and Life science (drug discovery, literature review)	EHR note, simulation of molecular interactions, summary of biomedical texts	41
	S/W Eng. (code generation, debugging, HDL, design, privacy analysis)	LLM (Copilot, CodeLLaMA), Programming, H/W logic synthesis	42
Enterprise & Business	Finance & Banking (Fraud detection, Chatbot, Reporting)	Transaction analysis, power financial assistant, automation of compliance summary	43-45
	Manufacturing & Supply (forecasting, logic analysis, training)	Forecasting of demand, Interpret logic, LLM-based tutor for engineering education	
	Legal & Regulatory (legal search, contract, compliance monitoring)	CoCounsel. Harvey AI for legal reasoning, risk detection	
Creation & Social domain	Creative industry (writing, Art/Music, Design idea)	LLM (MuseNet), Power story generation, Composition of music, Architecture sketches	
	Education (Conversation, Tutoring, Engament)	Support inclusive, always available learning	
	Training (Content customization, Feedback, Real-time assistance)	Platform AI (Khanmigo, Duolingo) for learning experience, skill development	
Autonomous system	LLM Agent (Task training, API interaction, Digital autonomous)	Auto-GPT, LangChain, for reason tool automation of workflow	46

Hochan Lee et al. [7] research that the impact of AI assistant on foreign rmployees' job performance and satisfaction in South Korean context. This study presents the potential of LLM-based AI assistants to alleviate labor shortages in East Asia, with a focus on South Korea by attracting foreign workers. This study reveals that the advantages of AI assistants are particularly pronounced among foreign employees. All participants noted that AI assistants significantly improved communication and translation, leading to enhanced job performance and adaptability. The findings suggest that AI was beneficial across all levels of Korean language proficiency, with those possessing lower proficiency relying on it more heavily. Its research does not mention about labor replacement by ChatGPT.

Guilherme C. Souza [8] analysis that GAI (Generative AI) for financial analysis, which is specifically large language models (LLMs). He shows how financial analysts read, screen and interpret information. He surveys practical uses already appearing in daily workflows for summarizing regulatory filings, analyzing earnings-call transcripts, building thematic baskets, tracking news and social-media sentiment and drafting research notes.

Qiang Chen et al. of ref. [9] show that AI master econometrics as AI Agent on expert-level tasks. They have a question that AI can effectively perform complex econometric analysis traditionally requiring human expertise or not? This paper evaluates an agentic AI's capability to master econometrics, focusing on empirical analysis performance. They develop a module called an Econometrics AI Agent built on the open-source MetaGPT framework. They illustrate that this agent exhibits outstanding performance in: (1) planning econometric tasks strategically, (2) generating and executing code, (3) employing error-based reflection for improved robustness, and (4) allowing iterative refinement through multi-round conversations.

They construct two datasets from academic coursework materials and published research papers to evaluate performance against real-world challenges. Comparative testing shows our domain-specialized agent significantly outperforms both benchmark large language models (LLMs) and general-purpose AI agents. This work establishes a testbed for exploring AI's impact on social science research and enables cost-effective integration of domain expertise, making advanced econometric methods accessible to users with minimal coding expertise. Furthermore, our agent enhances research reproducibility and offers promising pedagogical applications for econometrics teaching.

Ref. [10] illustrates that integrating traditional technical analysis with multi-agent LLM-based stock market forecasting. This paper introduces a multi-agent system that integrates the Elliott Wave Principle with AI for stock market forecasting. The inherent complexity of financial markets, characterized by non-linear dynamics, noise, and susceptibility to unpredictable external factors. It means that it has much significant challenges for accurate prediction. To show these challenges, this paper employs LLMs to enhance natural language understanding and decision-making capabilities within a multi-agent framework. By using technologies such as RAG (Retrieval-Augmented Generation) and DRL (Deep Reinforcement Learning), ElliottAgents performs continuous, multi-faceted analysis of market data to identify wave patterns and predict future price movements. The research results in the system's ability to process historical stock data and the results recognize Elliott wave patterns and generate actionable insights for traders. The experimental results of this paper perform on historical data from major U.S. companies, validate the system's effectiveness in pattern recognition and trend forecasting across various time frames. This paper illustrates that they contribute to the field of AI-driven financial analysis by demonstrating on how traditional technical analysis methods are effectively combined with modern AI approaches to create more reliable and interpretable market prediction systems.

Ref. [12] study that the rise of generative AI for modelling exposure, substitution, and inequality effects on the US labor market.

McKinsey & Company [13] the economic potential of GAI but it is not a analysis paper of economic trend and prediction by LLM.

As Korean economic analysis, Ref. [13] Pre-study of AI-based modelling and research for exact prediction of Korean economic trend. The first aim of his paper is to provide study strategies and simulation on how AI-based generative model and related technologies apply to economic analysis processing and what we have to prepare and study Korean econometrics. However, this paper does not provide the detailed research data because of pre-study.

III. LLM AND ECONOMETRICS RESEARCH

All paragraphs must be indented. All paragraphs must be justified, i.e. both left-justified and right-justified.

A. Overview for Study of Econometrics

Econometrics refers to the economic theory and statistical techniques for the purpose of testing hypothesis, estimating, and forecasting economic phenomenon. Literally, econometrics means economic measurement. Therefore, econometrics deals with the measurement about economic relationships of economic theory with mathematics and statistics. Prof. Ragnar Frisch of a Norwegian economist and statistician, first of all named this science as "Econometrics" in 1926. Econometrics is for studying economics phenomena. It is one of science which combines economic theory with economic statistics and tries by mathematical and statistical methods to investigate the empirical support of general economic law established by economic theory.

All studying econometrics is so wide and we cannot study once a time all topics. Therefore, econometrics is concerned with applying statistical and mathematical tools to economic data in order to give empirical content to economic theories. The scope of econometrics refers to the range of areas, activities, and applications where econometric techniques are used. It shows how econometrics connects economic theory with real-world data such as testing economic theories (Testing whether demand falls when price rises, Verifying the Phillips Curve: inflation vs unemployment, Checking if money supply affects GDP growth, Estimation of Economic Relationships (Estimating the price elasticity of demand, Measuring the marginal propensity to consume (MPC), Estimating the impact of education on wages), Forecasting Economic Variables, Forecasting GDP, inflation, unemployment, exports, interest rates, Predicting agricultural yields or tax revenue), Policy Formulation and Evaluation (Evaluating whether fiscal stimulus increases GDP, Measuring the effect of subsidies, scholarship schemes, or MGNREGA outcomes, Assessing the impact of GST on tax revenue, econometrics supports evidence-based policymaking), Testing Economic Hypotheses: Statistical Inference), Testing if the coefficient of price in a demand function is statistically significant, Testing whether returns on education are different for males and females), Studying Structural Changes (Whether the consumption function changed after liberalization, Whether export responsiveness changed after GST, Detecting breaks before and after economic crises or policy reforms, Dealing with Real-World Economic Problems (Real-world data often suffers from issues like, Multicollinearity, Heteroscedasticity, Autocorrelation, Simultaneity and Measurement error), Building Empirical Models (Econometrics develops models that describe real-life behavior and predict future trends, Consumption models, Investment functions, Demand & supply models, Production functions, Labor market models, Growth models and Financial econometric models), and Business and Financial Applications (Price forecasting for products, Predicting sales, demand, and market share, Stock market and financial return modelling, Risk measurement: e.g., Value at Risk using time series models).

B. Linear Regress Model (LRM) or Econometrics

Fig. 6 shows the domain of econometrics. In this paper, we use LRM (the Linear Regression Model) method to compare with ChatGPT.

The LRM is one of the most important tools in econometrics. It shows the relationship between a dependent variable (Y) and one or more independent variables (X). It helps us understand how changes in X affect Y [36]. Linear regression explains how a dependent variable can be expressed as a linear function of one or more independent variables as shown in equation (1).

$$Y = a + bX + u \quad (a = \text{intercept}, b = \text{slope (effect of X on Y)}, u = \text{error term}) \quad (1)$$

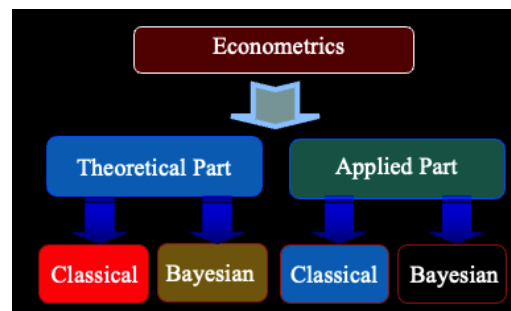


Fig. 6 The research domain of econometrics

The meaning of equation (1) in econometrics:

Dependent and Independent Variables

- Dependent variable (Y): the variable we want to explain (e.g., income, demand).
- Independent variable (X): the variable that influences Y (e.g., education, price).

Intercept and Slope

- Intercept (a): value of Y when X = 0.
- Slope (b): how much Y changes when X increases by one unit. Example: If b=0.5, then a one-unit increase in X raises Y by 0.5 units.

Error Term (u)

The error term represents all other factors affecting Y that are not included in the model. It reflects: measurement errors, omitted variables and random shocks.

Simple vs. Multiple Regression

- Simple Linear Regression: one X variable.
 $Y = a + bX + u$

- Multiple Linear Regression: more than one X variable. $Y = a + b_1X_1 + b_2X_2 + \dots + u$

Estimation Using OLS

The model is usually estimated using the Ordinary Least Squares (OLS) method. OLS chooses values of a and b such that the sum of squared errors is minimum. This gives the best-fitting line through the data.

Assumptions of Linear Regression (Basic)

Some basic assumptions include:

- Relationship between X and Y is linear
- Error term has zero mean
- Error term has constant variance
- No perfect multicollinearity (in multiple regression) ensure that OLS estimates are reliable.

Uses of Linear Regression

- Forecasting
- Measuring relationships between variables
- Testing economic theories
- Policy analysis

ChatGPT answer for PCA [OpenAI, ChatGPT]:

What PCA can (and can't) do for GDP forecasting

- **PCA is not a forecasting model by itself.** It only extracts a few common “global growth factors” from many GDP series.
- Forecasting happens **after** PCA: you forecast the **PC scores over time** (e.g., with VAR / AR / ARIMA / state-space), then reconstruct GDP forecasts.

Recommended workflow (practical)

1) Choose data + transform

Use a panel like:

- annual real GDP growth (%), or
- quarterly real GDP (log) → then growth rates ($\Delta \log$)

Typical transforms:

- Convert to growth: $g_{it} = 100 * (\log(\text{GDP}_{it}) - \log(\text{GDP}_{i,t-1}))$
- Standardize each country series (important for PCA): z-score per country.

2) Handle missing data

Real GDP panels often have missing entries.

Options:

- Keep balanced sample (drop countries with many gaps) — simplest
- Fill gaps with interpolation / EM-PCA / softImpute — more advanced

3) PCA on the standardized panel

Let \mathbf{X} be a $T \times N$ matrix (T time, N countries/regions). PCA gives:

- **Loadings** \mathbf{W} ($N \times K$): how each country loads on global factors
- **Scores** \mathbf{F} ($T \times K$): the time series of global factors (PC1, PC2, ...)

Usually:

- **PC1** \approx “global growth cycle”
- **PC2** \approx “EM vs DM divergence” or “commodity cycle” (often)

4) Forecast the factor(s)

Fit a time-series model to \mathbf{F} :

- **AR(p)** for each factor (fast, robust)
- **VAR(p)** for multiple factors (captures interactions)
- **ARIMA / ETS** if you want richer univariate dynamics
- **State-space / Kalman** if you want a full dynamic factor model

5) Reconstruct GDP forecasts

If you standardized \mathbf{X} before PCA, reconstruction is:

- $\hat{\mathbf{X}}_{\text{future}} = \mathbf{F}_{\text{future}} @ \mathbf{W.T}$
Then un-standardize per country (reverse z-score), and if you modeled growth, optionally integrate back

Basically, the LRM (Linear Regression Model) provides a simple form to understand. However, it is a powerful way to study economic relationships. It also helps quantify the effect of one variable on another and is the foundation of modern econometric analysis. The scope of econometrics is extensive and growing. It covers testing theories, estimating relationships, forecasting, policy evaluation, and solving real-world data problems. By connecting theory with data, econometrics plays a central role in modern economic analysis and decision-making.

C. Preparation for ChatGPT

This paper prepares some reference papers to compare with other results such as PCA (Principal Component Analysis), Deep learning, and Machine Learning for forecast of global GDP trend and Korean GDP trend.

The Forecast of Global GDP Trend

It shows the Python code and copy into Google colab.

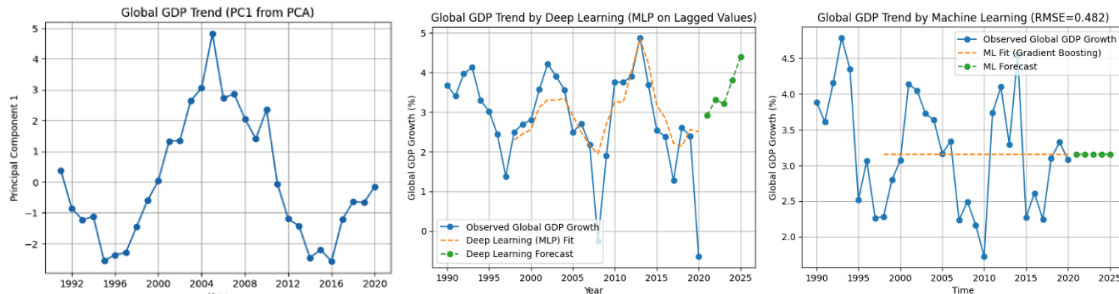


Fig. 7 The forecast of global GDP trend by PCA, DL, ML of ChatGPT

ChatGPT answer for machine learning [OpenAI, ChatGPT]:

Global GDP Trend by machine learning

You can do “Global GDP Trend by machine learning” by turning the GDP growth time series into a supervised learning problem:

- Inputs = **lagged values** (last L years/quarters), optionally rolling stats
- Model = **RandomForest / GradientBoosting / XGBoost / LightGBM**
- Output = next-step GDP growth
- Then plot **observed vs ML fit + forecast**

Below is a **working, fast scikit-learn version** (no deep learning).

The forecast of Korea GDP Trend by ChatGPT

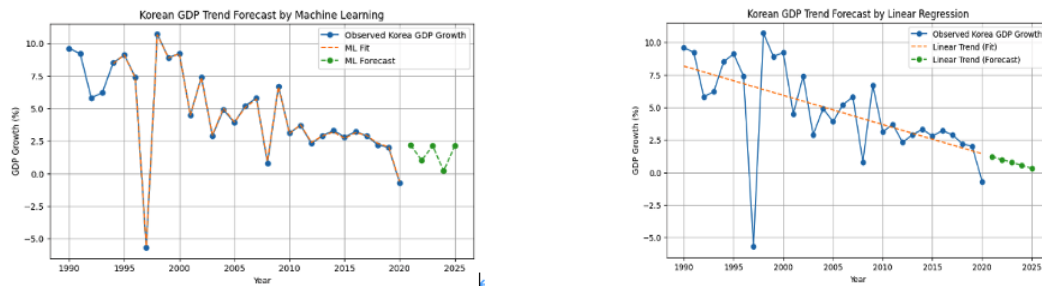


Fig. 8 The forecast of global GDP trend by DL and LR of ChatGPT

The equation of ChatGPT for forecast

$$GDP\ Growth = -0.2244 + Year_t + 454.71$$

For example, for 2025:

$$GDP\ Growth_{2025} \approx 8.2 - 0.2244 \times 35 \approx 0.4\%$$

D. Research Strategy of K-EcoLLM Model

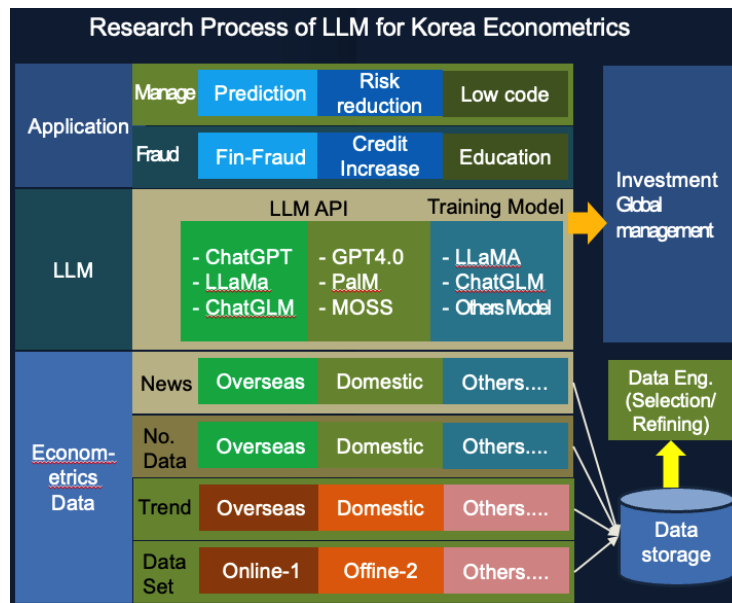


Fig. 9 The research process of K-EcoLLM Model

Fig. 9 shows the research strategy of K-EcoLLM (Korea Econometric LLM). The main key of research is to apply to large language models (LLMs) of econometrics domain. This group is composed of Data Source, Data Engineering, LLM, and Applications.

- **Data source layer:** This is the Data Source Layer is the acquisition of extensive financial data from a wide array of online sources. This layer ensures comprehensive market coverage by integrating data from news websites, social media platforms, financial statements, market trends, and more. The goal is to capture every nuance of the market, thereby addressing the inherent temporal sensitivity of financial data.
- **Data engineering layer:** This layer focuses on the real-time processing of data in financial situation.
- **LLM layer:** It deals with the highly dynamic nature of financial data and ensure its responses with the current financial situation.
- **Application layer:** The final component of K-EcoLLM is the applications layer, which is designed to demonstrate the practical applicability. It always provides tutorials and demo of applications for financial tasks, advisory services, quantitative trading, investment information, information for management, forecasting of GDP growth, and so on.

E. Discussion for Application of LLM

Large language models (LLMs) and ChatGPT is one of machines with function but it should be trained by correct data for economic research. They are AI systems trained to predict the next word given preceding text, and are typically fine-tuned to follow human instructions and generate responses. The performance of LLMs improves predictably according to scaling factors that are empirical regularities and data.

LLM is subject to significant data limitations. The pretraining process uses data up to a certain point in time, and the knowledge of the LLM cuts off at that point.

Despite these limitations, LLMs exhibit such quick response times and low transaction cost that they are useful for automating a wide range of micro-tasks in which they are still error prone and in which similarly capable human research assistants would not be competitive. the instantaneous response of LLMs makes it useful to outsource this task, even if there are occasional mistakes. Similarly, I would not hire a human research assistant who regularly commits basic logical fallacies while presenting results with great confidence.

LLMs are quite capable at coding tasks, especially in the languages Python and R, for which the most online resources are available. LLMs can write, edit, modify, translate, or debug snippets of code based on instructions in plain English (or other natural languages).

Moreover, they can act as tutors when using new libraries, functions, or even programming languages that the user is not very familiar with by quickly generating out-put that shows what libraries and functions are needed for specific types of operations or what syntactic structures to use in a given programming language. This allows the human programmer to consult the LLM and weave together code from many different snippets generated by it. The reasons LLMs are so proficient at coding include the following: There are vast repositories of code available online that are included in their training data, for example from GitHub. The syntax of computer code seems to be relatively easy to learn for these systems. Moreover, the AI labs producing cutting-edge LLMs themselves benefit from the code generation abilities of these systems, which provides them with strong incentives for improving these capabilities. In fact, coding may be one of the areas where current LLMs lead to the greatest productivity gains: Peng et al. (2023) report a controlled experiment in which programmers using OpenAI-powered GitHub Copilot completed their assignments on average 55.8 percent faster, amounting to a 126 percent productivity increase.

Fig. 7 and Fig. 8, and equations of above should be analyzed on how much it is close and true. To understand these questions, we should study again through by other tools and descriptions. Here, this paper shows the possibility and LLM results of econometrics. And then this paper suggests a strong necessity for research of Korean econometrics.

ChatGPT Advanced Data Analysis

To use LLM effectively, they must suggest the result of data analysis about economic issues they study domain and topic. Of course, if they show the simulation results, it will be better. With this, reader can understand and believe LLM function.

Simulating Human Subjects

The power of LLM is to present human subjects through directly communication that they require or obtain it. Through this observation about subject, the training data of LLMs contain a large amount of information about humanity for their focuses. They condition GPT-3 on the sociodemographic backstories of real humans and demonstrate that subsequent answers to survey questions are highly correlated with the actual responses of humans with the described backgrounds in a nuanced and multifaceted manner.

Deriving Equations

To effective research, they should show their results by equation. At the process of writing, LLMs also should illustrate as the example of equation. LLMs still routinely make mistakes in mathematical derivations when users enter. The LLM process to re-derive the results, corrected the mistake but current LLM make a mistake or possess the high-level abstract representations that they should necessarily perform mathematic.

Explaining Models

Current LLMs also have some ability to explain simple models. This is useful, but also risky for students. That is why we must research on our topic.

IV. CONCLUSION

This paper suggests the motivation and the study strategy of Korean econometric by LLM (K-EcoLLM). Nowadays, the possibility of LLM application is wider and deeper in science and engineering as well as general purpose because of model ability such as text reading, image recognition function, translation ability of image and text, including art generation. Herein, market of LLM increases more for economy. Several report and research papers predict that generative AI will change higher labour productivity, higher education, and higher. It means the role of LLM is bigger and bigger and will be super generative and influence economic growth factors. Therefore, this paper provides material for prediction of economy trend and motivation to research Korean econometrics by LLM (K-EcoLLM). Unfortunately, we cannot use the LLM that develop for the general purpose because the situation of economic growth is so complex and dynamically change in real-time.

As shown in Table 1, LLM (Large language model) have a strong necessary for economists to predict from many materials because of easy communication between bothers. That is why they should offer application of LLM design, training, adaptation, evaluation, and application for both researchers and practitioners for Korean econometric research. Unfortunately, in Korea, there are few application materials of LLM in econometrics. To accomplish this, this paper provides the necessary material for user that they are interested in econometrics by LLM. The research of LLM should mention the evolution of language models, core approaches, PEFT (Parameter-Efficient Fine-tuning), and practical data for application through research process. The methodology of LLM application should mention the survey of real-world LLM applications across the scientific area, engineering field, healthcare issues, and creative topic and method through a review of current Korean economic situation.

**REFERENCES**

- [1]. <https://en.wikipedia.org/wiki/OpenAI>
- [2]. Brady D. Lund, "A Brief Review of ChatGPT: Its Value and the Underlying GPT Technology", University of North Texas. DOI:10.13140/RG.2.2.28474.06087
- [3]. Brady D. Lund and Ting Wang, "Chatting about ChatGPT: How may AI and GPT impact academia and libraries?", DOI: 10.1108/LHTN-01-2023-0009
- [4]. Dinesh Katta, "Study and Analysis of Chat GPT and its Impact on Different Fields of Study", international Journal of Innovative Science and Research Technology Vol. 8, Issue 3, March – 2023, Colorado Technical University.
- [5]. Allison Slater Tate, "How will AI like ChatGPT change education for our children", <https://www.parents.com/how-will-ai-technology-change-education-7100688>.
- [6]. Tianyu et al., "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development," IEEE/CAA, Vol. 19, No. 5, 2023.
- [7]. https://bold.expert/technology/?filter-category%5B%5D=education-technology-technology&gclid=Cj0KCQjww4-hBhCtARIsAC9gR3aJCWHu0LzYNBGqGoZ6A1lb6Lb2y-6f—lhdiBSV1UJeaon3ID_bcIaAnj9EALw_wcB
- [8]. Matt Zalaznick, "5 ways ChatGPT will driver deeper learning instead of more cheating. <https://districtadministration.com/5-ways-chatgpt-will-drive-deeper-learning-instead-of-more-cheating/>
- [9]. Dong Hwa Kim, "How to teach and Learn AI. OutskirtPress", USA, 2023.
- [10]. Liam Frady, "Chat GPT-4 vs Chat GPT-3: What's the Difference, and Which Is Better? 2023.
- [11]. Arianna Johnson, Forbes Staff, "Bard Vs. ChatGPT, The Major Differce Between The AI Chat Tools, 2023.
- [12]. Sanuj Bhatia, "OpenAI GPT-4 Features, Comparison with ChatGPT, and How to Use It", 2023.
- [13]. (Partha Pratim Ray,, "ChatGPT, A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet of Things and Cyber-Physical Systems, 2023.
- [17]. Marcus et al., "The economic impact of generative AI", The future of work in India, June 2023.
- [18]. <https://www.anthropic.com/news/claude-3-5-sonnet>.
- [19]. Michal et al., "Integrating Traditional Technical Analysis with AI: A Multi-Agent LLM-Based Approach to Stock Market Forecasting", DOI: 10.5220/0013191200003890.
- [20]. <https://hls.harvard.edu/graduate-program/ll-m-program/>
- [21]. <https://law.yale.edu/studying-law-yale/degree-programs/graduate-programs/llm-program>
- [22]. <https://hls.harvard.edu/graduate-program/ll-m-program/>
- [23]. <https://www.postgraduate.study.cam.ac.uk/courses/directory/lwlwllll>
- [24]. <https://www.law.uchicago.edu/llm>
- [25]. <https://law.illinois.edu/academics/degree-programs/llm-program/>
- [26]. Wel Zgao et al. "Road of Large Language Model: Source, Challenge, and Future Perspectives", <https://doi.org/10.34133/research.0655>, 2025.
- [27]. Steffen et al. "Transforming Science with Large Language Models: A Survey on AI-assisted Scientific Discovery, Experimentation, Content Generation, and Evaluation", <https://chatgpt.com/>, <https://github.com/QwenLM/Qwen>, <https://www.deepseek.com/>, 2025.
- [28]. Dong Hwa Kim, "How to Educate AI Thinking, AI Literacy, and AI Literature?", IARJSET. Vol. 12, No. 9, 2025.
- [29]. Dong Hwa Kim, "Impact on Job Replacement and Economy by TAG-AI", IARJSET, Vol. 11, No. 1, 2024.
- [30]. Dong Hwa Kim, "A Study on the Impact of Generative AI on Korean Economy Growth", IARJSET, Vol. 10, No. 9, 2023.
- [31]. Dong Hwa Kim et al. "Pre-study of AI-based Modelling and Research for Exact Prediction of Korean Economic Trend", IARJSET, Vol. 13, No. 1, 2026.
- [32]. Marcel Blnz. "How should the advancement of large language models affect the practice of science?", PNAS, Vol. 122, 2025.
- [33]. Junhyuk Choi, "Can LLM Simulate Economics Experiment with 522 Real-human Persona?", 2025.
- [34]. Kyoungseo Lee, Forecasting Korea's GDP growth rate based on the dynamic factor model", KJAS, Vol. 37, No. 2, 2024.
- [35]. Junxi Wu. "Modeling and Forecasting Global GDP Trends Using PCA and Machine Learning: Evidence from 1960 to 2020", [h□ps://dl.acm.org/doi/10.1145/3746972.3746983](https://dl.acm.org/doi/10.1145/3746972.3746983), 2025.
- [36]. https://en.wikipedia.org/wiki/Linear_regression.
- [37]. Jatin et al. "A Comparative Study of Large Language Models ChatGPT DeepSeek Claude and Qwen", DOI: 10.1109/DICCT64131.2025.10986449, 2025.
- [38]. Mostafa, "Generative AI, Large Language Models, and ChatGPT in Construction Education, Training, and Practice", Buildings, 15, 2025. <https://doi.org/10.3390/buildings15060933>.

- [39]. Julien, et al. "LLMs vs. econometric models for nowcasting GDP growth: A practitioner's view", <https://hal.science/hal-05128225v1>, 2025.
- [40]. Eger, S. et al. "Transforming Science with Large Language Models: A Survey on AI-assisted Scientific Discovery, Experimentation, Content Generation, and Evaluation", arXiv preprint arXiv:2502.05151 2025.
- [41]. Mohammadabadi, et al. "Examination of AI's role in Diagnosis, Treatment, and Patient care: In Transforming Gender-Based Healthcare with AI and Machine Learning", CRC Press, pp. 221–238, 2024.
- [42]. Mohammadabadi, et al. "Towards distributed learning of pmu data: A federated learning based event classification approach", In Proceedings of the 2023 IEEE Power & Energy Society General Meeting (PESGM). IEEE, pp. 1–5, 2023.
- [43]. Li, Y., et al. "H. Large language models in finance: A survey. In Proceedings of the Proceedings of the fourth ACM international conference on AI in finance", pp. 374–382, 2023.
- [44]. Peddinti, S.R., et al. "A. Utilizing large language models for advanced service management: Potential applications and operational challenges", Journal of Science & Technology Vol. 4, 2023.
- [45]. Lopez-Lira, et al. "Bridging language models and financial analysis", arXiv preprint arXiv:2503.22693 2025.
- [46]. Yang, H. et al. "Auto-gpt for online decision making: Benchmarks and additional opinions", arXiv preprint arXiv:2306.02224 2023.

BIOGRAPHY



Dong Hwa Kim Ph.D.: Dept. of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering (AI Application for Automatic control), TIT (Tokyo Institute of Technology), Tokyo, Japan. He worked at the Hanbat National University (Dean, Prof., S. Korea); Prof. at Electrical Power and Control Eng. Adama Science and Tech. Uni., Ethiopia; TDTU, Vietnam. He has experience in many universities overseas as Prof. He was NCP of EU-FP7 (EU-Framework Program, ICT). He had a keynote speaker at several international conferences and universities. He has 200 papers in journals and conferences. He is reviewing IEEE and other's journals. He is currently a researcher at the Seoul national university of S&T. He published many books and papers such as Innovation tuning based on biotechnology (USA, Dec. 2017), 4th wave Status and preparation of Visegrad Group Country (Germany, 2019), How to They Education in the Famous Univ. (2019), Africa and 4th Wave: Will it risk or Chance? (Amazon, 2020), How to teach and Learn AI (Outskirt Press, USA, Aug. 2022), A Study on Reinforcement of Self-Directed Learning Using Controlling Face Emotion (Paper, Jan. 2022), Advanced Lectures for PID Controller of Nonlinear System in Python (IJRTE, March 2021), Dynamic Decoupling and Intelligent Optimal PID Controller Tuning Multivariable Qua-drones (IJRTE (Scopus), Dec. 2021), Failure Prediction of Wind Turbine using Neural Network and Operation Signal (IJRTE, Dec. 2021), and 200 papers.

- Home page: www.worldhumancare.wixsite.com/kimsite

- Research citations: https://www.researchgate.net/profile/Dong_Kim53Current, DSTSC. Executive.