

서울 R
Meetup

챗GPT와 데이터 사이언스

챗GPT와 7월

Seoul R Meetup

ChatGPT and its Futures



Dr. Kim, Song-Kyoo (Amang)
Associate Professor,

Faculty of Applied Sciences
MACAO POLYTECHNIC UNIVERSITY
Macau, SAR





Agenda

- About (Chat) GPT
 - What is the (Chat) GPT?
 - Challenges of (Chat) GPT
 - Future development plan
- (Chat) GPT Impacts
 - Academia
 - Industry (Job Market)
 - Culture & Trends
- Post GPT Era
 - How to prepare the post-GPT Era
 - Future Prediction*
 - Geocentrism & (Chat) GPT

What is ChatGPT ? (1/2)

Depends on Hallucination (h)

$$\lim_{h \rightarrow 0} GPT(h) = \textit{Google Search}$$

$$\lim_{h \rightarrow \infty} GPT(h) = \textit{Flirter (Charlatan*)}$$

$$\lim_{h \rightarrow h^*} GPT(h) =$$



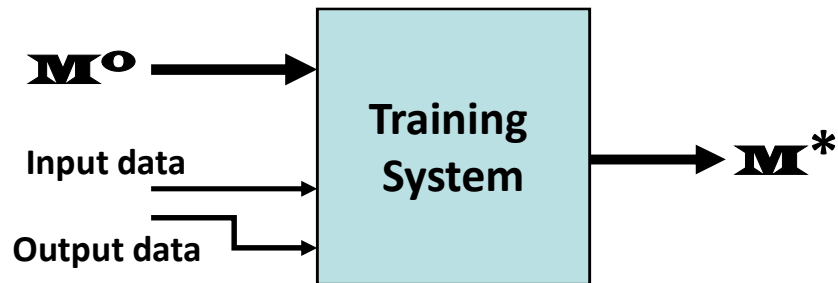
**) Perfect liar, "The best lie is the one that's hidden inside the truth."*



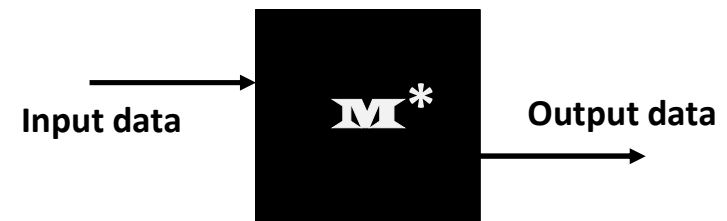
What is ChatGPT (2/2)

● Conventional ML System

Training

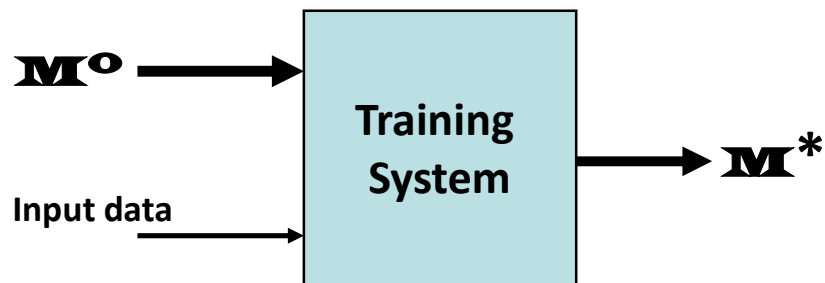


Testing/Inference

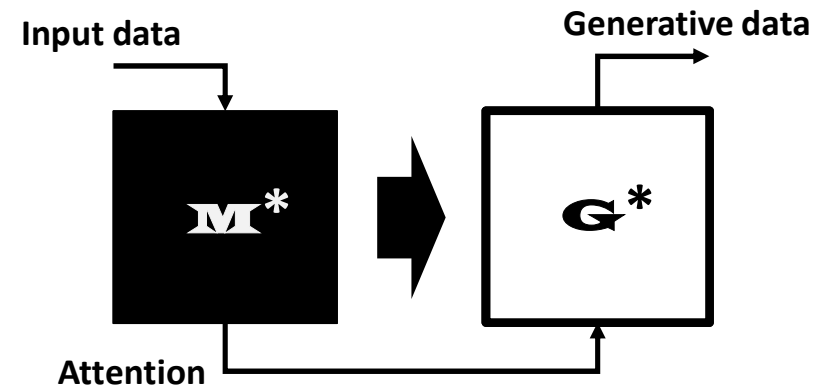


● Generative ML System

Training



Testing/Inference





Challenges of (Chat) GPT (1/4)

● Hallucination

- Degree of freedom (df)
- $df \rightarrow \infty$; Full generator
- $df \rightarrow 0$; Analyzer (searcher) (i.e., $G^*(df) = I$)

● Harm of representation, allocation, quality of services.

- Continue to reinforce social biases and worldviews
- Contexts of high risk government decision making
 - Law enforcement, criminal justice;
 - Migration and asylum;
 - Offering legal or health advice.



Challenges of (Chat) GPT (2/4)

● Disinformation & Influence Operations

- Misinformation (fake news);
- Propaganda (여론선동)

● Harmful contents

- Advance or encourage self-harm;
- Graphical materials (e.g., XXX, violence);
- Harassing, hate contents;
- Planning attacks or violent;
- Illegal contents;

● Overreliance (과다 의존성)



Challenges of (Chat) GPT (3/4)

- **Economic impacts by policy makers**
 - Job changes;
 - Workforce displacement (인간대체)
- **Cybersecurity**
 - Vulnerability discovery (네트워크 취약점 발견);
 - Social engineering (Phishing);
- **Privacy**
 - Trained by public information then combining several information to generate (or reasoning) private information.



Challenges of (Chat) GPT (4/4)

- **Proliferation** of conventional & unconventional weapons
 - Faking likelihood confidential information.
 - Proliferation: 확산
- **Potential risks for emergent behaviors**
 - emergent: short thinking
 - emergent deception by intention (i.e., majority, mean) → Human involvement (RLHF, RBRM).
- **Acceleration** (가속화)
 - Lose small but vital things.



Future Development of GPT

- RLHF (Reinforcement Learning Human Factor).
- RBRM (Rule Based Reward Model).
- User intent by RLHF.
- Closed domain hallucination.
- System safety.
 - Usage policies and monitoring;
 - Contents classifier development;

Major Limitations (태생적 한계)

The major limitation of GPT could be found on its name...

GPT:

Generative Pre-trained Transformer

“Generative” + “Pre-Trained”



GPT Impacts on Academe (1/6)

● Impacts on Secondary and Higher Educations

- Public schools in New York city have banned to use AI tools:
 - ChatGPT could generate very convincing responses and essays in response to user prompts.
- Essay writing is one of most commonly used assessment tool to assess student's ability of critical-thinking.
 - Although it is the fact that logical-thinking is not critical-thinking, people majoring humanity and/or social science still myth that logical thinking skill could be raised by critical thinking.
- Simple essay assignments without logical (or scientific or mathematical) reasoning are no longer proper assessment to check critical-thinking skill.
 - University admission procedures (US) shall be expected to be modified for exactly the same reason.



GPT Impacts on Academe (3/6)

● Impacts on Research Scholars

- The increasing popularity of AI-based tools in research and writing has led to a growing interest in exploring the potential benefits of ChatGPT in these fields.
- There has been significant debate surrounding the ethical considerations associated with AI-chatbot authorship.
- ChatGPT holds the potential to offer substantial benefits to researchers whose native language is not English, particularly those from Asian regions such as China, Japan, and Korea.
- ChatGPT is even capable to generate a comprehensive literature review of a specified length based on a set of provided references.
- ChatGPT has the capability to generate a wide range of qualitative documents.



GPT Impacts on Academe (4/6)

● Writing examples by using ChatGPT

TABLE I. CHATGPT ASISSTED EDITING EXAMPLE

Prompt: Revise the following paragraph to resarch tone:
Many students are already using ChatGPT for their take-home assignments. ChatGPT seems more suitable for doing remote exams although they might not have good grades but they will pass most courses in business, management, law and social sciences. It is noted that ChatGPT provides mediocre performances within those subjects.
ChatGPT:
More and more students are turning to ChatGPT for their homework assignments and it has proven to be an effective solution for remote exams in disciplines such as business, management, law, and social sciences. Although top grades may not always be achieved, students have a good chance of passing the majority of their courses in these fields. It's important to note that ChatGPT's performance in these subjects is considered to be standard.

TABLE II. ChatGPT asissted Literature Review

Prompt: Provide 10 international journal article references for artificial intelligence with IEEE reference format.
ChatGPT:
[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[2] Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.
[3] V. Nair and G. Hinton, "Rectified linear units improve restricted boltzmann machines," in Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.
...
[9] Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.
[10] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.



GPT Impacts on Academe (5/6)

● Practical strategies for facing challenges

■ Plagiarism of assignments:

- Teachers should be aware of the possibility of students using ChatGPT and other AI applications to complete their assignments.
- Although simple plagiarism checks may not be foolproof, they can still be effective in detecting AI-generated content.
- As technology continues to advance, it is expected that the ability to detect AI-generated content will also improve.
- Short-term strategy:
 - Essay + the follow up questionnaire based on their essay.
 - Unique and (or) not-institutive scientific questions.

■ Intellectual Properties and Copyrights:

- AI applications shall be simple tools for improving our daily lives.
- Most research work today requires the use of computers or specific software packages, yet the names of these technologies are not listed as authors in research papers.

GPT Impacts on Academe (6/6)



- Technologies are growing for both good and bad sides.

The Lending Club dataset contains 887,379 loan records, and the authors removed the loans whose status was "current", leaving 269,668 loan records for analysis, as described in Figure 7. Figure 8 depicts the status of all loans in the dataset. The authors split the loan data into two categories: default and full payment, with a ratio of about 3.5:1 in their dataset.

The authors experimented with different sampling methods on three machine learning algorithms: random forests, neural networks, and logistic regression. They use different resampling methods, including random undersampling, SMOTE, Tomek link, and boundary SMOTE, to balance the dataset. Logistic regression with resampling or cost-sensitive learning outperforms random forests and neural networks. The best sampling method is random undersampling because it achieves the highest F1 score compared to other sampling methods. The authors also found that the optimal alpha values for logistic regression with cost-sensitive learning were 0, 3.2, and 3.4 based on accuracy and accuracy default values, respectively. The authors selected the top ten significant features resulting from performing random forests, which provided the best performance. They also tried using only the first three important features and found that the results were not significantly different from those using all features.

Match Overview		
37%		
1	Yen-Ru Chen, Jenq-Shi... Publication	23%
2	doaj.org Internet Source	2%
3	www.cakeresume.com Internet Source	2%
4	Soreti Bekele Babo, Asr... Publication	2%

Introduction

The document highlights the rapid growth of P2P lending in recent years as an alternative way for small businesses and individuals without credit histories to obtain credit without involving financial institutions such as banks. However, the fundamental problem of P2P lending is information asymmetry, which may not be able to accurately estimate the default risk of lending. Datasets in P2P lending are often unbalanced, containing unequal fully paid and defaulted loans. The paper highlights the challenges of dealing with imbalanced datasets in the context of P2P lending, proposes the use of resampling and cost-sensitive mechanisms to deal with imbalanced data. The paper also utilizes logistic regression, random forest, and neural network algorithms to predict default risk, and validates the proposed scheme using the Lending Club dataset.

- Section 2: This section reviews related work on predicting P2P lending default risk and classifying imbalanced datasets. The authors emphasize the importance of addressing the problem of imbalanced datasets in P2P lending default risk classification, and discuss various techniques for dealing with imbalanced data, such as resampling, cost-sensitive learning, and ensemble learning.
- Section 3: This section describes the methodology used in the study, including the datasets used, preprocessing steps, and the machine learning algorithms used to predict default risk. The authors explain the use of under-sampling and cost-sensitive learning to deal with imbalanced datasets and describe the feature selection and feature engineering techniques employed. Finally, they detail the logistic regression, random forest, and neural network algorithms used to predict default risk.
- Section 4: This section shows the evaluation metrics and experimental results of the proposed scheme. The authors compare the performance of three machine learning algorithms and demonstrate the effectiveness of

Page 1 of 5

How much of this submission has been generated by AI? **30%**

of qualifying text in this submission has been determined to be generated by AI.

How do we detect AI-generated writing?

To learn more about Turnitin's AI writing detection model and how it works, please visit our [AI writing detection page](#).



GPT Impacts on Industry (2/6)

● Demolishing the value of contents creators

- Creative art works are no longer created by human only.

[제목 : 고양이 사라졌다]

캐서린 : "아웅~ 아웅~! 천천히 와볼까?"

(고양이 소리와 함께 캐서린이 화면 밖으로 나온다)

친구A : "이게 무슨 일이야? 왜 고양이가 울고 다녀?"

캐서린 : "고양이가 사라졌어. 방금까지까지 봤는데 갑자기 없어졌어."

친구B : "어디 갔어? 뭐라도 있었어?"

캐서린 : "아무것도 없어. 그냥 사라졌어."

(친구들이 둘러싸인 캐서린. 갑자기 문이 열린다.)

남자 : "미안하다. 내가 그 고양이를 데려갔어."

캐서린 : "뭐? 왜?"

남자 : "그냥 이유없이. 죄송해."

(남자가 고양이를 내려놓고 나서 빠르게 사라진다.)

친구A : "뭐야 이게? 대체 뭐지?"

캐서린 : "나도 몰라. 그냥 이상하게 끝난 것 같아."

(화면이 어두워지면서 끝남)

이렇게 작성해드렸는데, 원하시는 다른 형식이 있다면 말씀해주세요!

3 minutes drama scripts

To create a 1-minute smooth jazz guitar piece using the given chord progression, you'd need to expand and vary the melody, rhythms, and techniques. Here's an idea of how you might structure the 1-minute piece using the 2-5-1 progression:

Chord progression:
||: Dm7 | G7 | Cmaj7 | A7 :|| Repeat this 4 times

Tablature notation (melody):

Section A:
E|-----6-----6-----|-----8-----6---|---5---5---5/7|-----|
B|---6---6---6---6---|---8---8---8---|---5---5---7|---8---7---|
G|---7---5---7---7---|---7---7---|---5---4---|---6---|
D|-----|-----|-----|-----|
A|-----|-----|-----|-----|
E|-----|-----|-----|-----|

Section B:
E|-----8-----8-----|-----10-----8---|---9---9---9/11|-----|
B|---8---8---8---|---10---10---|---9---11---|---12---11---|
G|---9---7---9---9---|---9---9---9---|---9---8---|---10---|
D|-----|-----|-----|-----|
A|-----|-----|-----|-----|
E|-----|-----|-----|-----|

Structure: A A B A

Play Section A 2 times, then Section B once, and Section A once for a total of about 1 minute.

Note that this is a simplified and basic concept. To fully develop a 1-minute piece,

1 minute jazz guitar codes



GPT Impacts on Industry (3/6)

- Demolishing the value of consulting and story telling which is based on the collective intelligence.

Strengths	Weaknesses
<ul style="list-style-type: none">• Global brand recognition and reputation• Diverse product portfolio• Strong research and development capabilities• Vertical integration• Market leadership	<ul style="list-style-type: none">• High reliance on the smartphone market• Legal issues• Limited software capabilities
Opportunities	Threats
<ul style="list-style-type: none">• Expansion into emerging markets• Investments in advanced technologies• Collaborations and partnerships• Focus on eco-friendly and smart home products	<ul style="list-style-type: none">• Intense competition• Price-sensitive markets• Rapid technological changes• Geopolitical and economic risks

SWOT Analysis for Samsung's Future Strategy



GPT Impacts on Industry (4/6)

● Which jobs (or industry) shall be terminated?

■ Contents creators / Artists

- Illustrators/painters, music composers, ...
- Webtooniest, Youtuber, ...

■ Business consulting

- business, management, marketing, brand, ...
- New product/business/service development, ...

■ Knowledge Coordinators (Consulting)

- Story tellers (e.g., Youtuber) based conventional knowledge.
- Various coordinators including fashion, beauty, law; medical; real estimate; ...



GPT Impacts on Industry (5/6)

● Which jobs shall be terminated? (cont.)

■ Writers / Speakers

- Commercial copy writers, reporters,
- Scripters (drama, speech, ...)
- Creative writings (manuscripts, novels, ...)

■ Planner (travel, marriage, ...)

■ Programmer / coder

■ Customer service representative

- telephone/e-mail/(chat) messengers, ...



GPT Impacts on Industry (6/6)

● Polarization

- These jobs are NOT be terminated but polarized.
- Hyper competitive red ocean and extremely rare blue ocean are coexisted.
- Getting more polarized until the red ocean is bust (i.e., easy and (almost) free to use).
- As long as the values exist, the jobs are never disappeared (blue ocean but getting more rare).
- The job description shall be redefined.



GPT Impacts on Culture & Trend (1/4)

● Demolishing the value of artworks

- Even elementary school kids could make decent pieces of artworks.



Draw “Macao”



GPT Impacts on Culture & Trend (2/4)

- Demolishing the value of the collective intelligence (i.e., teamwork leadership)



Using ChatGPT

SWOT Analysis for Samsung Electronics' Strategy for 2023

Strengths <ul style="list-style-type: none">Very high research and development (R&D) expenditure resulting in one of the strongest patent portfolios among technology companiesIndustry leadership in most of the markets the company operates, such as smartphones, TVs, memory chips and home appliancesBrand recognition and reputation among consumers and business partnersEnormous advertising and marketing budget to promote its products and servicesProduct innovation and design capabilities that cater to different customer segments and preferencesVertical integration and manufacturing outsourcing that result in low production costs and high quality standards	Weaknesses <ul style="list-style-type: none">Too many smartphone models that confuse customers and dilute the brand imagePositioning the company as a high-tier, mid-tier and low-tier smartphone vendor that reduces its profitability and premium statusDependence on Google's Android OS that provides little control over the user experience and exposes the company to potential legal issuesHigh competition from other technology companies, especially from Chinese rivals like Huawei, Xiaomi and Oppo that offer cheaper and comparable productsDeclining market share in some emerging markets due to political tensions, regulatory barriers and consumer preferences
Opportunities <ul style="list-style-type: none">Expanding into new markets and segments, such as cloud computing, artificial intelligence, 5G, IoT and wearable devicesLeveraging its strong patent portfolio and R&D capabilities to develop new products and services that meet the changing customer needs and expectationsStrengthening its online presence and e-commerce channels to reach more customers and increase salesBuilding strategic partnerships and alliances with other technology companies, such as Microsoft, Google and Amazon, to enhance its product offerings and competitive edgeInvesting in corporate social responsibility and sustainability initiatives to improve its brand image and reputation among stakeholders	Threats <ul style="list-style-type: none">Increasing competition from other technology companies, especially from Chinese rivals that offer cheaper and comparable productsRegulatory changes and legal disputes that affect its operations and profitability, such as antitrust investigations, patent infringements and trade warsCybersecurity risks and data breaches that compromise its customer data and trustTechnological disruptions and innovations that render its products and services obsolete or less attractiveEconomic downturns and social unrest that reduce the consumer demand and spending on its products and services



GPT Impacts on Culture & Trend (3/4)

- Before the (mechanical) computer...
 - **Mathematics == (simple) Calculation**
- After the computer...
 - **Mathematics != (simple) Calculation**
- Before AlphaGo (Analytical ML)
 - **Go is only for human** because it is unlimited choices and too many uncertainties (which is not true).
- After AlphaGo
 - **Go is no longer only for human** because it is limited choices and the uncertainty can be calculated.



GPT Impacts on Culture & Trend (4/4)

● Before ChatGPT (Generative ML)...

- Writing new things, making artworks (illustration), composing new music are representing “Creativity” and only human can do these works.
- Working together is mandatory for developing the recommendations from various perspectives (i.e., collective intelligence).

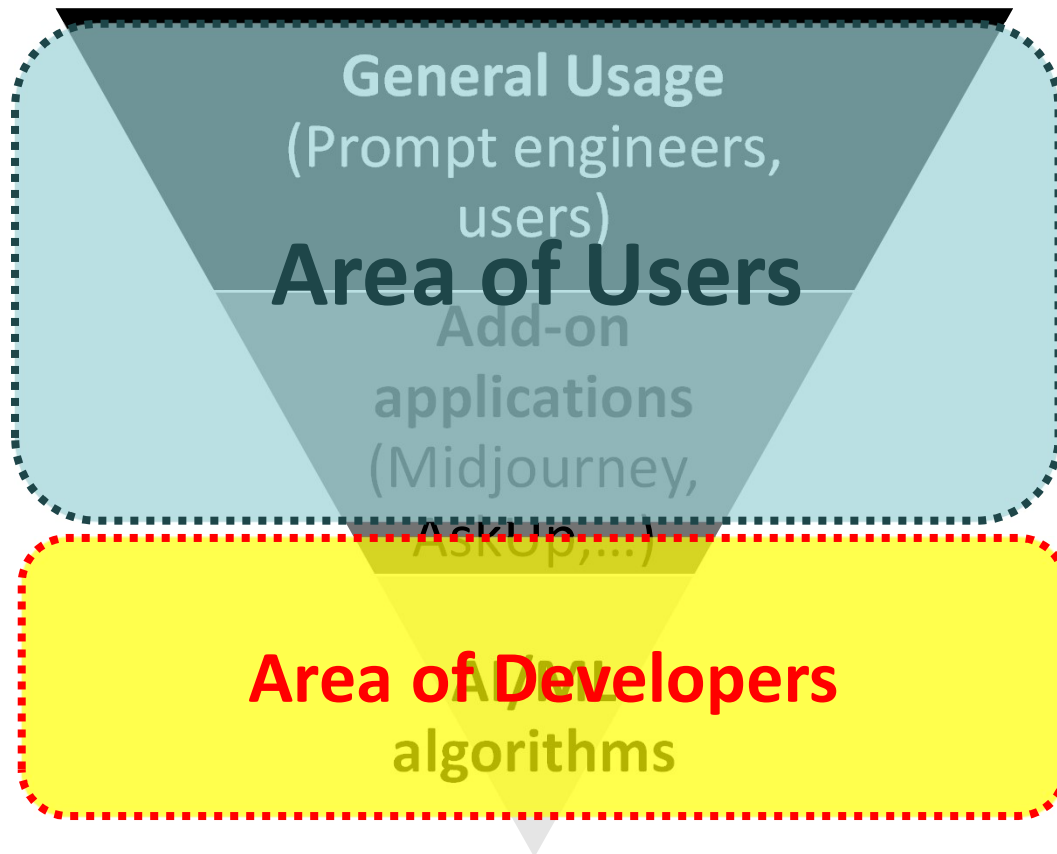
● After ChatGPT...

- Machine could do decent creative works (even kids could make quality artworks and literatures).
- No teamwork is required to get better recommendations from various perspectives.



How to prepare the post-GPT Era (2/5)

● The Position to Learn the ChatGPT



① For better using current existing AI/ML models.



② Developing add-on applications on the top of existing AI/ML models.



③ Developing AI/ML algorithms (LLM, GAN, CNN, ...)



How to prepare the post-GPT Era (3/5)

● What should you learn to prepare the future?

■ Area of Users:

Liberal Arts + Domain Knowledge

■ Area of Developers:

Liberal Arts + Specialized CS & Math & Physics #

● Liberal Arts

■ **Liberal Arts** = Languages + Sciences + Humanities + Arts;

■ Languages = Literature* (English) + Mathematics + Coding**

■ Sciences = Physics^{##} + Chemistry + Biology

■ Humanities = History + Sociology + Economics^{##} + Ethics

■ Arts (skills) = Art (visual) + Music + PE

■ In Korean → 국영, 수, 과, 사탐



How to prepare the post-GPT Era (5/5)

- Actually, you do not need to learn the ChatGPT.
 - If you don't come to the generative MLs (e.g., Chat GPT), they will come to you.
- ChatGPT Add-ons

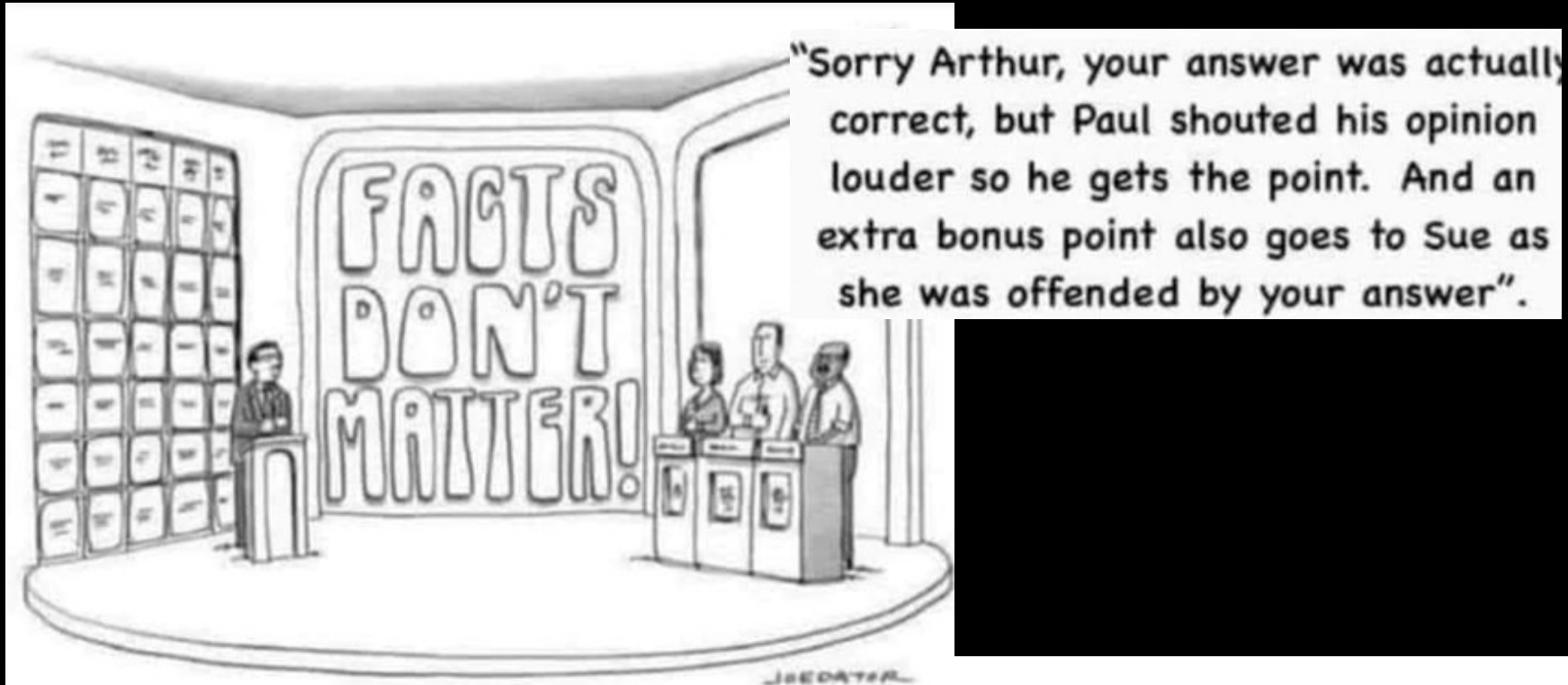
Video	Muse AI	Visla AI	Topaz AI
Text	ChatGPT	Notion AI	Compose
Images	Midjourney	Magic Studio	Pebblely
Design	Viesus	Piggy AI	Galileo

Coding	Bugasura	CodeGPT	Replit Ghostwriter
Audio	FineShare	Boomy AI	Playlist AI
Productivity	Briefly AI	Socra AI	Leexi AI

(Source: Anonymous, Internet)

Geocentrism & GPT (1/2)

Future Jeopardy (Quiz Show)



죄송합니다, 아서(Arthur). 당신의 답변은 실제로 정확했지만, 폴(Paul)이 더 큰 소리로 자신의 의견을 외쳤기 때문에 점수를 받게 되었습니다. 그리고 수(Sue)도 작성자의 답변에 분개하여 보너스 점수가 추가로 부여되었습니다.

(Source: Internet)

Geocentrism & GPT (2/2)

What if the (Chat) GPT has been invented on the era of the Geocentrism?

Heliocentrism would never been revealed !!



Overall Summary

- The ChatGPT is not a science, it is a humanity.
- Redefine the meaning of “Creativity”
 - Most creativity which currently defined are no longer considered as creativity.
 - “Generative” is no longer “Creative”
- If you don’t come to technologies, they will come to you.
- Liberal Arts are vital for surviving the post GPT era.
- As usual, only Basics are survived but they could not be the star in the business.

