Medical Imaging & Intelligent Reality Lab.
Convergence Medicine/Radiology

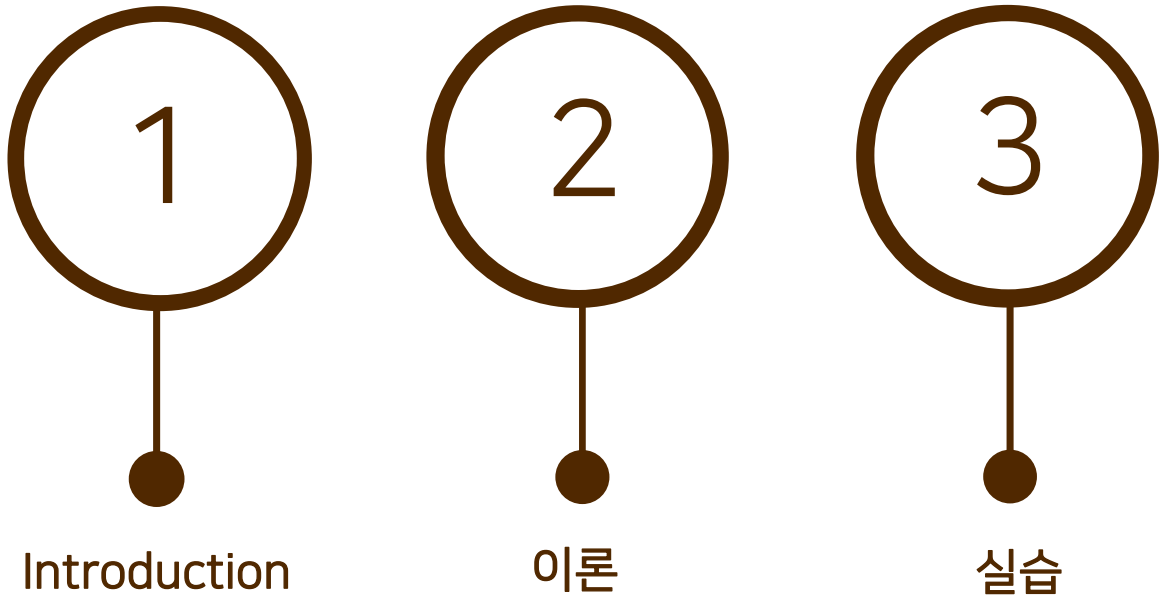
ChatGPT를 이용한 의료인공지능 개발

Git: https://github.com/babbu3682/Med_tutorial_ChatGPT

Presenter: Sunggu Kyung

Email: babbu3682@gmail.com

Program Contents

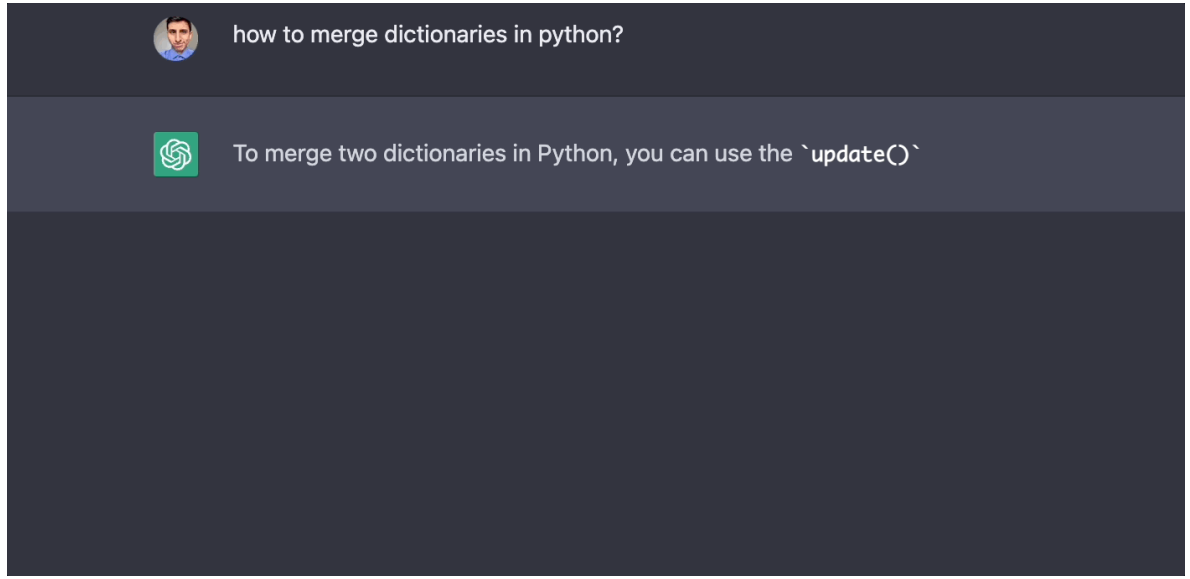


ChatGPT

- ChatGPT

- ChatGPT는 Instruct GPT의 형제 모델로, 프롬프트의 instruction을 따르고 자세한 응답을 제공하도록 교육 되었습니다.

- ChatGPT는 OpenAI에서 개발한 GPT 기반의 **대화형** 인공지능 모델입니다.



 # https://openai.com/blog/chatgpt

ChatGPT

장점:
- 넓은 범위의 코드 응답을 생성할 수 있으며 코드 개념을 설명하는 데 도움이 된다.
- 비기술적 이해관계자에게 더 적합하며 더 유연합니다: **사용자는 거의 모든 질문에 답할 수 있는 도우미와 채팅 같은 인터페이스를 통해 소통합니다.**
- 이 유연성은 사용자가 질문을 명확히 하고 재구성할 수 있게 해서, 더 정교한 코드를 추출할 수 있다.

한계:
- 복잡하거나 기술적인 언어에 어려움을 겪을 수 있으며, **항상 정확하거나 관련 있는 응답을 생성하지 않을 수 있다.**
- 게임이나 거래와 같은 실시간 상호작용이 필요한 애플리케이션에는 적합하지 않을 수 있다.

 # https://www.techtarget.com/searchenterpriseai/tip/GitHub-Copilot-vs-ChatGPT-How-do-they-compare

유용한 ChatGPT 확장프로그램

1. 프롬프트 입력

- **프롬프트 지니: 한국인들을 위한 실시간 번역 앱.**
- **AIPRM for ChatGPT: 원하는 형태의 답변을 미리 세팅해준다. (강추)**

2. 결과물 응용 작업

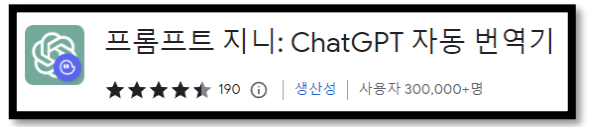
- ShareGPT: 결과물을 외부와 공유하고 파일로 다운로드 할 수 있다.
- Table Capture: 테이블 형식의 결과를 복사 및 다운로드할 수 있습니다.

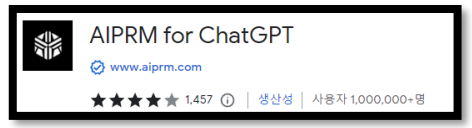
3. 웹 검색 참조

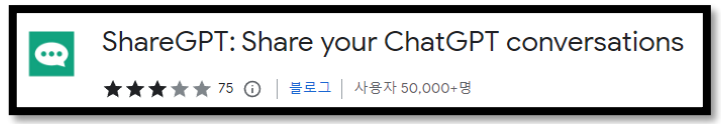
- WebChatGPT: 검색 결과를 ChatGPT에 반영하고 출처 링크를 표시해 줍니다.

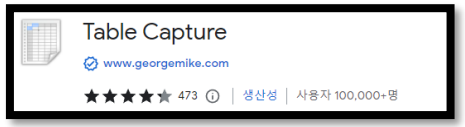
4. 검색 엔진

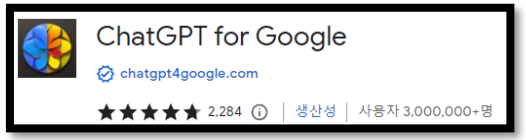
- ChatGPT for Google: Bing처럼 크롬에서도 ChatGPT를 이용하는 방법.











Prompt Engineering

Guidelines for Prompting

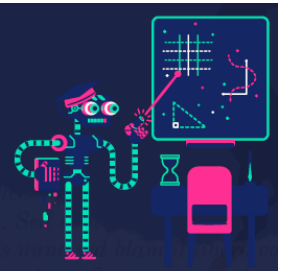
Data Exploration

You collect the data. Now it is time to explore it!

Prompts

1. What is Data Exploration, and why is it important in Data Analysis?
2. What are some common techniques used in Data Exploration?
3. What is the purpose of Data Cleaning, and what are some common data cleaning techniques?
4. How do you handle missing values in a dataset during Data Exploration?
5. What are some common ways to identify outliers in a dataset?
6. What is the purpose of Data transformation, and what are some common data transformation techniques?
7. How do you check for and handle multicollinearity in a dataset?
8. How do you perform feature scaling in a dataset, and what are some common scaling techniques?
9. What is the purpose of Data Visualization in Data Exploration?
10. What are some common types of charts and graphs used in Data Visualization?
11. How do you create a scatter plot in Python, and what information can you gain from it?
12. What is a histogram, and how is it useful in Data Exploration?
13. How do you create a box plot in Python, and what information can you gain from it?
14. What is a heatmap, and how can it be used to explore relationships in a dataset?
15. How do you create a line plot in Python, and what information can you gain from it?
16. What is the purpose of Exploratory Data Analysis (EDA), and how is it performed in Python?
17. What are some common statistical measures used in EDA, such as mean, median, and mode?
18. How do you use correlation analysis to explore relationships between variables in a dataset?
19. What is the purpose of hypothesis testing in Data Exploration, and what are some common tests used in Python?
20. How do you use Data Exploration to identify potential areas for further analysis or research?

Machine Learning

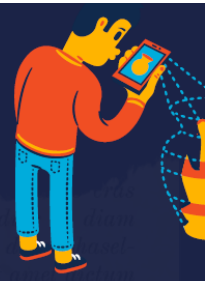


You got the data and you discover the data. Now it is time to create values by making predictions with Machine Learning!

Prompts

1. What is the difference between Linear and Logistic regression?
2. What evaluation metrics can be used to evaluate the performance of a Regression model?
3. How can you handle multicollinearity in a Regression model?
4. What are some common techniques for feature selection in Regression models?
5. What is the difference between binary and multiclass classification?
6. How can you handle imbalanced datasets in classification problems?
7. What are the different evaluation metrics for Classification models, and when should each be used?
8. How can you interpret the results of a Classification model using techniques like feature importance or permutation importance?
9. What is the difference between hierarchical and partition clustering algorithms?
10. How can you determine the optimal number of clusters in a Clustering Model?
11. What are some common techniques for preprocessing data before Clustering?
12. How can you visualize the results of a Clustering Model?
13. What is tokenization, and why is it important in NLP?
14. What is the difference between stemming and lemmatization?
15. How can you vectorize text data for use in Machine learning Models?
16. What are some common techniques for preprocessing text data before using it in NLP models?
17. What is the difference between object detection and object recognition in computer vision?
18. How can you train a deep learning model for image classification using TensorFlow or Keras?
19. What is transfer learning, and how can it be used in computer vision?
20. How can you evaluate the performance of a computer vision model using metrics like accuracy, precision, and recall?

Data Visualization



You collect the data. Now it is time to explore it!

Prompts

1. What is Matplotlib and how is it used in Data visualization?
2. How do you create a scatter plot using Matplotlib?
3. How can you customize the style and formatting of a Matplotlib plot?
4. What are some common types of plots you can create using Matplotlib?
5. What is the difference between plt.plot() and plt.scatter() in Matplotlib?
6. What is Seaborn and how does it differ from Matplotlib?
7. How do you create a heatmap using Seaborn?
8. What are some common statistical plots you can create using Seaborn?
9. How can you customize the style and formatting of a Seaborn plot?
10. What is the difference between a heatmap and a clustermap in Seaborn?
11. What is Plotly and how is it used in data visualization?
12. How do you create an interactive line chart using Plotly?
13. What are some common types of interactive plots you can create using Plotly?
14. How can you customize the style and formatting of a Plotly plot?
15. How does Plotly differ from other Python visualization libraries?
16. What is Bokeh and how is it used in Data visualization?
17. How do you create an interactive scatter plot using Bokeh?
18. What are some common types of interactive plots you can create using Bokeh?
19. How can you customize the style and formatting of a Bokeh plot?
20. How does Bokeh support interactive visualization in web browsers?

 # https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson

Guidelines for Prompting

KDnuggets | Visit KDnuggets.com for more cheatsheets and additional learning resources.

ChatGPT Cheat Sheet

ChatGPT is a large language conversational AI built by OpenAI. It was trained using Reinforcement Learning from Human Feedback, similar to InstructGPT. ChatGPT understands the prompt and provides detailed repose that can help you with research, coding, and various data science tasks.

Ideas

Dataset Suggestion
>>> I want to build a predictive model for image classifiers. Can you please suggest the five most relevant datasets for my use case?

Suggest Resources
>>> I would like to learn about deep learning. Please suggest 3 best specific resources.

AB Testing
>>> [case-study]. Please design an A/B test for this purpose. Please include the concrete steps on which statistical test I should run.

Career Coaching
>>> I am looking for a role as a data engineer. My background is management. What should I do in 6 months to get a job?

Coding

Unit Test
>>> Write a unit test for train function. The test cases are: x should not be null value and y should be a numerical value.

Code Explanation
>>> Can you explain what the code is doing? `[code-snippet]`

Optimize Code
>>> Can you improve the time complexity of the code? `[code-snippet]`

SQL

SQL Formatting
>>> Format the following SQL code and convert all reserved keywords to uppercase. `[code-snippet]`

Translate Between DBMS
>>> What is the equivalent of PostgreSQL's DATE_TRUNC for MySQL?

Calculate Average
>>> Write the SQL code that works for PostgreSQL 14. I have a table with two columns [date, temp]. I would like to calculate an average temp.

Calculate Runway
>>> Write SQL to calculate my runway.

Spreadsheets

Spreadsheets Formula
>>> Create a spreadsheet formula to calculate the sum of cells B1 to B20?

Dummy Data
>>> Generate the dummy data for me to use as placeholders in my spreadsheet.

Tips
>>> Give me some tips on how to improve the efficiency of my spreadsheet?

Data Analysis

Generate Data
>>> Generate a fake data with 100 rows and 4 columns: [id,name,grade,subject]

Data Cleaning
>>> I have a text classification dataset. Write Python code for data cleaning.

Data Exploration
>>> I have a dataset of 100 rows and four columns:[id, name, grade, subject]. Write R code for data visualization and exploration.

Data Visualization
>>> I have a dataset with 100 rows columns [id, name, grade, subject]. Create a matplotlib bar chart of subject vs. grade.

Machine Learning

Train Regression Model
>>> You are a data scientist, write Python code for me. I have a dataset with columns [model,hp,speed]. Please build a machine learning model that predicts speed.

Hyperparameter Tuning
>>> I have a logistic regression model, write Python code to tune hyperparameters.

Imbalance Data
>>> I have an imbalanced dataset with target column species. In python, how do I oversample and/or undersample my data?

Explain the Model
>>> I have trained a LightGBM model. Write a Python code to explain the output using a series of plots with Shap.

Research

Explain the Concept
>>> Explain t-test to an undergraduate as a data science instructor.

Stakeholders
>>> Tips on explaining data science reports to a business stakeholder.

Summarize the paper
>>> Please summarize the paper "Adding Conditional Control to Text-to-Image Diffusion Models" in simple terms in one paragraph.

Writing Blog
>>> Write an outline for a blog "Python lists".

Research History
>>> Can you research the history of the graph neural network?

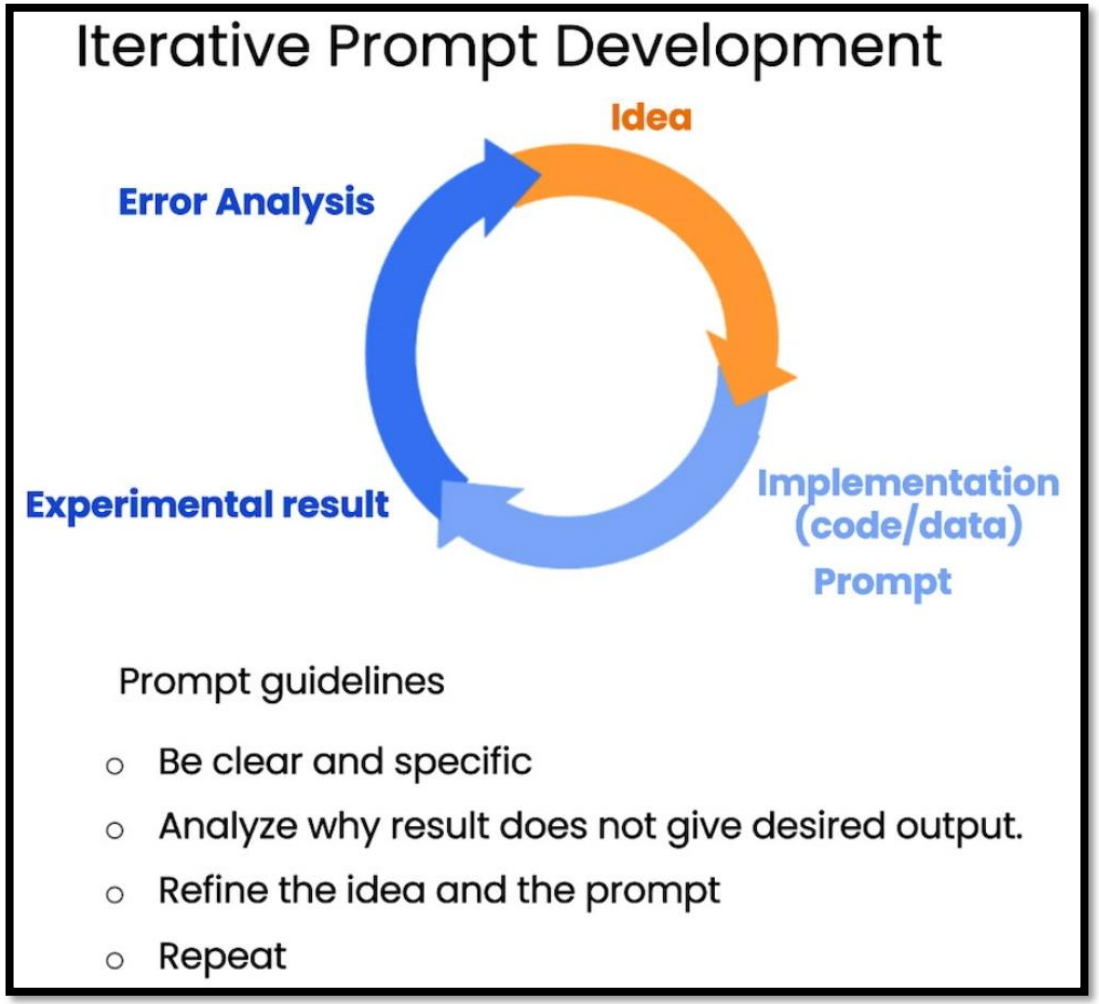
Subscribe to KDnuggets News

Abid Ali Awan | 2023

 # https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson

Iterative Prompt Development

➢ 완벽한 prompt는 없다. 반복된 과정으로 자신만의 prompt를 완성하라.



 # https://learn.deeplearning.ai/chatgpt-prompt-eng/lesson

Coding

Hands-On

Using Tools

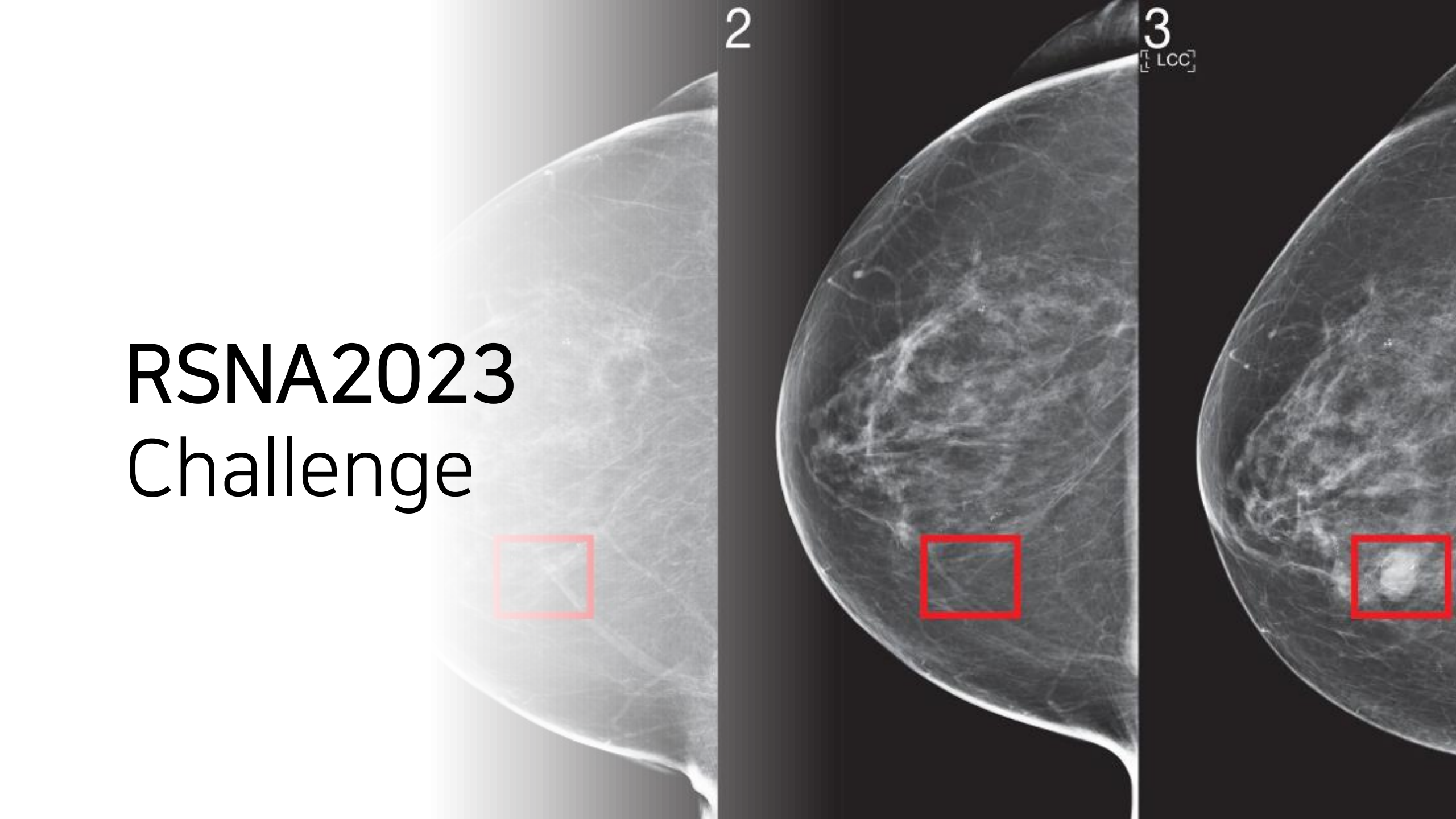
- AIPRM for ChatGPT

  - 참고: https://www.aiprm.com/blog/guidelines-how-to-write-public-prompt-templates-for-aiprm/

- 프롬프트 지니

  - 참고: https://www.promptgenie.ai/

RSNA2023
Challenge



AI Challenge

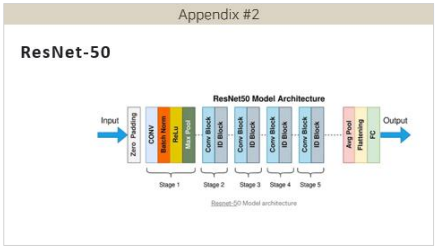
- 1. classification – RSNA2023

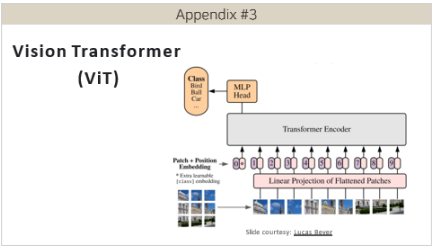
  → Mammography Breast Cancer Detection (Link)

- 강의를 위한 축소화 버전의 Dataset

Step-by-step

1. Base ResNet-50
2. ResNet-50 + Augmentation
3. ResNet-50 + Augmentation + weight decay
4. ViT-base + Augmentation + weight decay





RSNA Screening Mammography Breast Cancer Detection

- Challenge 목표:
  여성 대상 유방 방사선 이미지를 사용하며, 이 대회의 목표는 스크리닝 검사에서의 유방암 증례를 유방 촬영에서 식별하는 것입니다.

  유방암 식별 → Binary classification task

- Challenge 필요성:
  유방암은 세계적으로 가장 흔한 암이며, 조기 발견과 치료가 중요하다. 현재 유방암의 조기 발견 및 치료에는 고도로 훈련된 전문가의 지식이 필요하여 비용이 많이 들며, 몇몇 국가의 Radiologist 의 부족이 이 문제를 악화시킬 가능성이 높습니다.

- Challenge 의의:
  이 Challenge에 참여하여 자동화 모델을 개발함으로써 조기 발견의 혜택을 더 많은 사람들에게 확대할 수 있으며, 나아가 세계적으로 유방암 사망률을 낮출 수 있다.

 # https://www.kaggle.com/competitions/rsna-breast-cancer-detection/overview

Entire development process

EDA Process

- 1. Analysis the given data file

  (example)

  Demography
- 2. Analysis the given image file

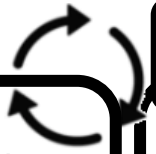
  (example)

  Histogram,

  Rescale Type,

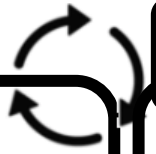
  Window center,

  Window width



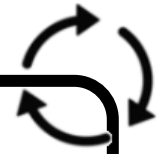
Training Process

- 0. Fix Seed
- 1. Dataset
- 2. Network
- 3. Mutli-GPU
- 4. Loss
- 5. Optimizer
- 6. Learning Rate (LR) scheduler
- (7. Resume)
- 8. Metric
- 9. Train & Valid Loop



Testing Process

- 0. Fix Seed
- 1. Log analysis
- 2. Dataset
- 3. Network
- 4. Using GPU
- 5. Loss
- 6. Resume
- 7. Metric
- 8. Test Loop
- 9. Result analysis



필수 라이브러리 - Pandas

- Pandas는 데이터 조작 및 분석에 사용되는 Python의 강력한 오픈 소스 라이브러리입니다.

- 사용하기 쉬운 데이터 구조 및 데이터 분석 도구를 제공하므로 구조화된 데이터 작업에 널리 사용됩니다.

- 판다스 10분 완성: https://dataitgirls2.github.io/10minutes2pandas/

 # https://www.kaggle.com/competitions/rsna-breast-cancer-detection/overview

필수 라이브러리 – Matplotlib, Seaborn

- Matplotlib는 Python에서 널리 사용되는 오픈 소스 도식화 라이브러리로 다양한 유형의 시각화를 생성할 수 있습니다.

- Seaborn은 Matplotlib을 기반으로 다양한 색상 테마와 통계용 차트 등의 기능을 추가한 시각화 패키지.

- Matplotlib 기초정리:
https://doorbw.tistory.com/173

- Seaborn 기초정리:
https://datascienceschool.net – 5장 데이터 시각화

 # https://www.kaggle.com/competitions/rsna-breast-cancer-detection/overview

RSNA Screening Mammography Breast Cancer Detection

- GitHub
- https://github.com/babbu3682/Med_ChatGPT_tutorial

 # https://www.kaggle.com/competitions/rsna-breast-cancer-detection/overview

Reference

- https://medium.com/@tanyamarleytsui/coding-with-chatgpt-b50ab3fcb45f
- https://thepromptartisan.com/prompt-engineering-in-chatgpt-a-comprehensive-master-course/

Thank you for your Attention….!

Appendix #0

axillary tail (AT)

craniocaudal (CC)

lateromedial (LM)

lateromedial oblique (LMO)

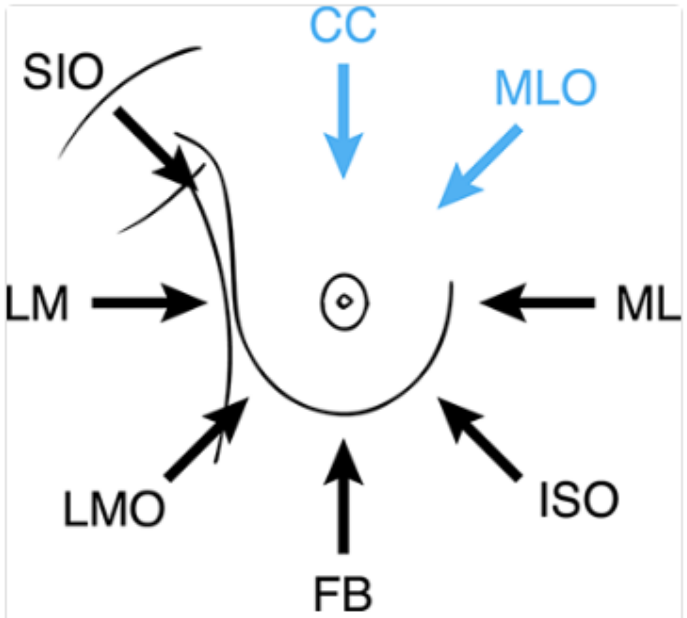
mediolateral (ML)

mediolateral oblique (MLO)

Digital mammography images

Our team was one of the best performes groups in the Breast Cancer Contest organized by DREAM.

Each screening mammography exam usually comprises two images per breast: a view from above called a craniocaudal (CC) view and an oblique or angled view called mediolateral oblique (MLO) view. These two views show the medial part as well the external lateral portion of the breast as much as possible. More than 99% of all the exams provided by Group Health have both CC and MLO views for each breast imaged. Figure 2 shows an example of CC and MLO views for the left and right breast of an healthy subject.



1 Breast orientation

It is important to note that the breast projection is very important, as this is a deformable organ, the main issue is to track about the location of the tumour.

http://med_physics.i-do.science/topics/diagnose_breast/

Intro to Explainability

Grad-CAM (Selvaraju et al.)

Integrated gradients (Sundararajan et al.)

Input X Gradient (Shrikumar et al.)

LIME (Ribeiro et al.)

KernelSHAP (Lundberg et al.)

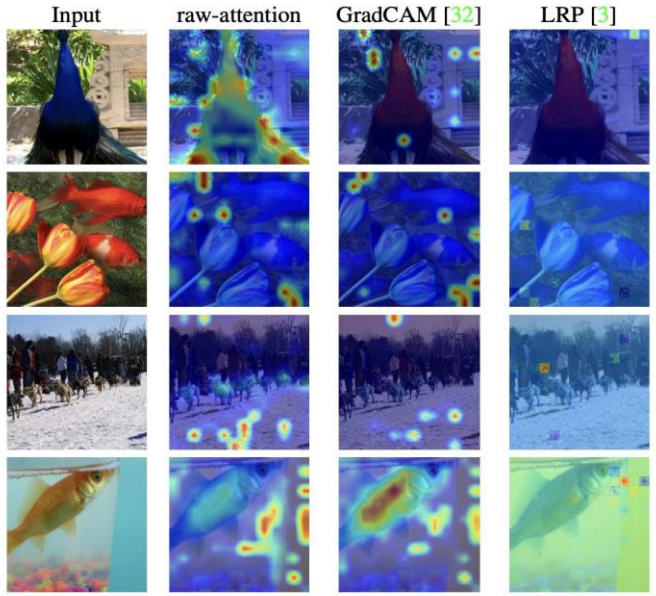
DeepLift (Shrikumar et al.)

And many more!

CNNs vs. Transformers

CNNs and Transformers differ significantly in the architecture

- Attention vs. convolution.
- For Transformers- **classification is mostly obtained by a CLS token.**



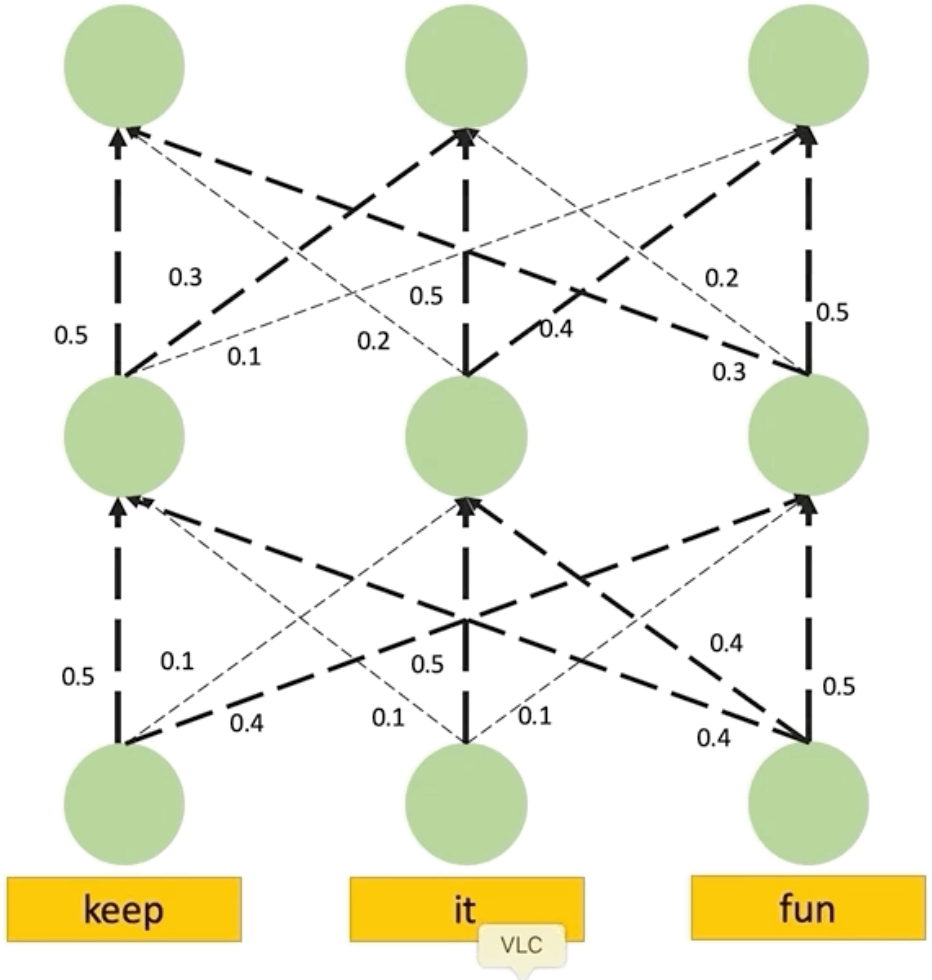
Transformer Interpretability Beyond Attention Visualization, Chefer et al., 2021

Solution #1- Attention Rollout

- Aggregation across heads: **averaging**.
- Aggregation across layers: **matrix multiplication of the attention maps to track context.**

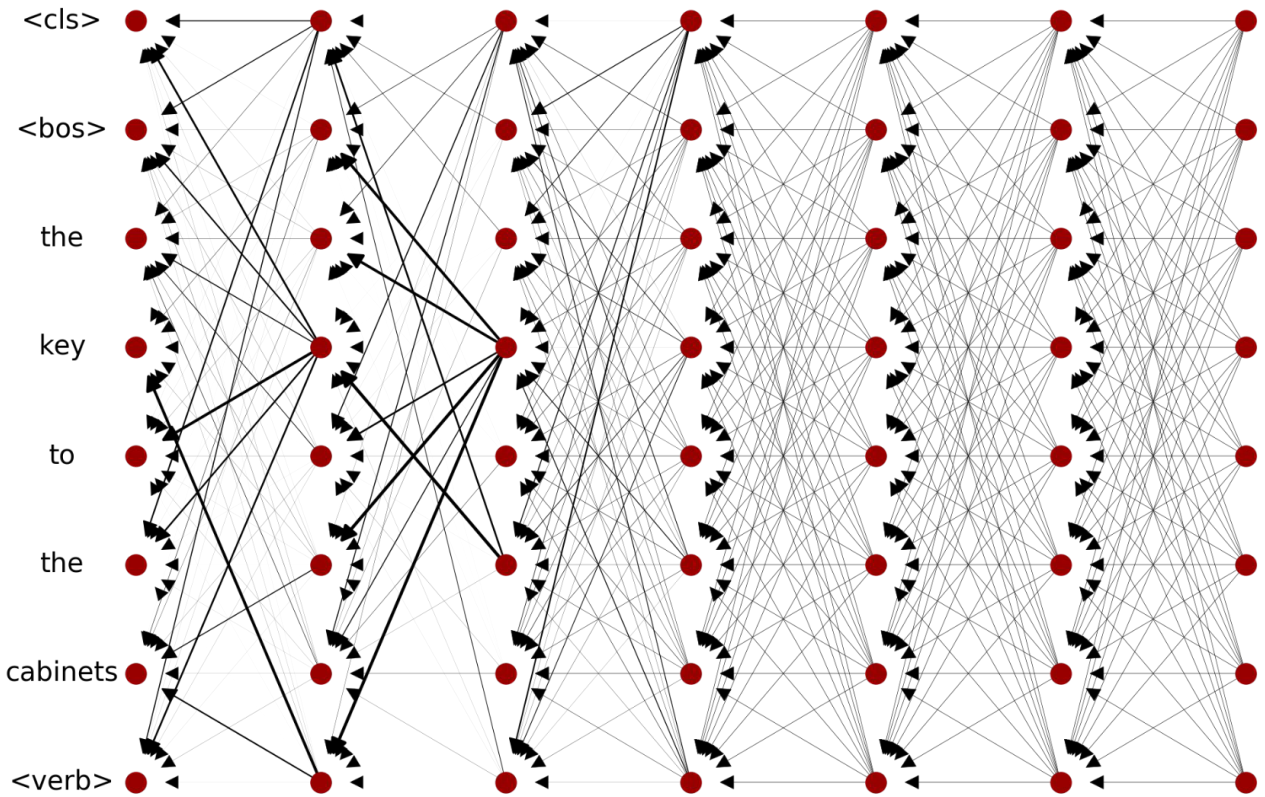
Quantifying Attention Flow in Transformers, Abnar et al., 2020

Solution #1- Attention Rollout



Quantifying Attention Flow in Transformers- Blog, Samira Abnar

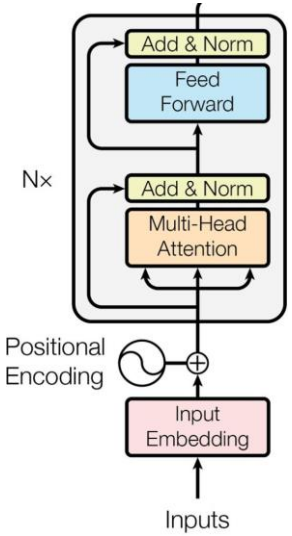
Solution #1- Attention Rollout



Quantifying Attention Flow in Transformers- Blog, Samira Abnar

Solution #1- Attention Rollout

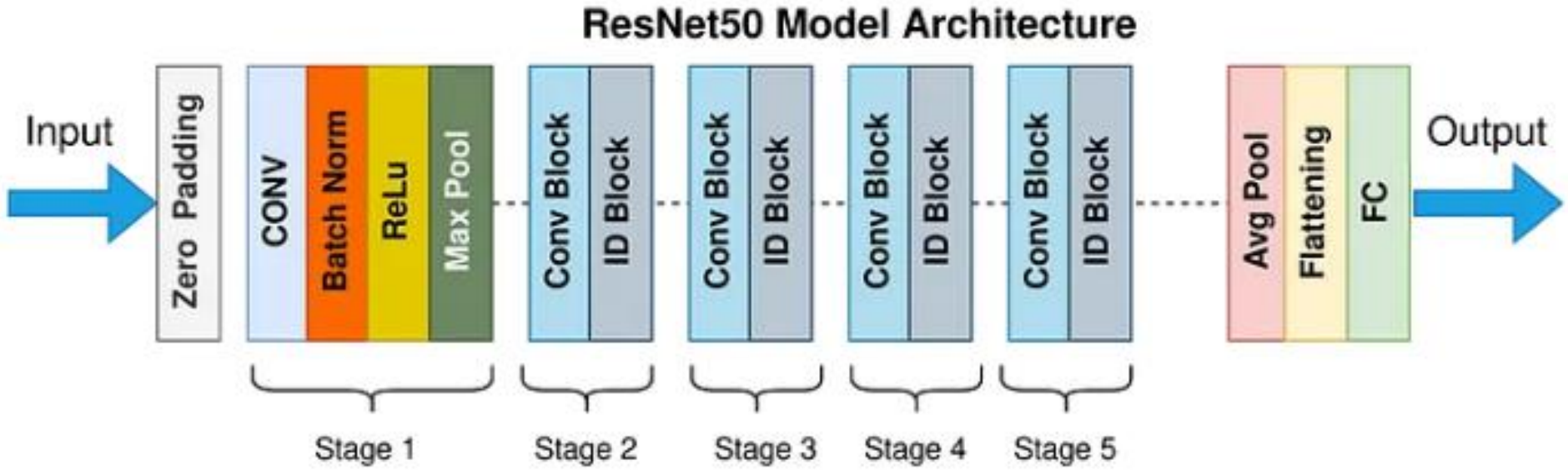
- Aggregation across layers: **matrix multiplication to track context:**
  - The attention maps of all layers are multiplied.
  - The identity matrix is added to each self-attention matrix to account for the residual connections.



$$A = 0.5W_{att} + 0.5I,$$

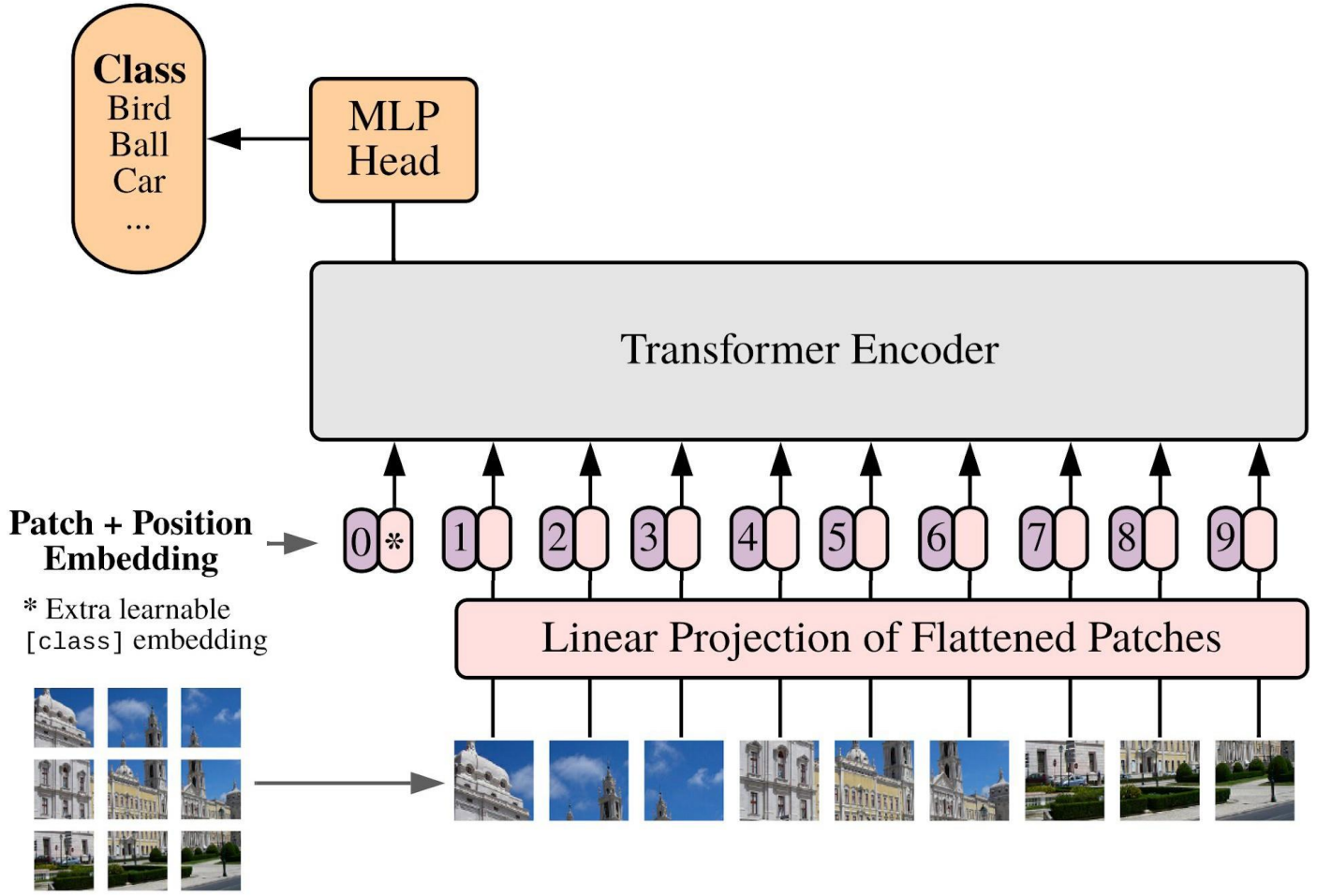
Quantifying Attention Flow in Transformers, Abnar et al., 2020
Image source: Attention is All you Need, Vaswani et al., 2017

ResNet-50



Resnet-50 Model architecture

Vision Transformer (ViT)



Slide courtesy: Lucas Beyer