SUPERCHARGE YOUR AI PIPELINE BY FINETUNING TRANSFORMER MODELS

DEEP DIVE INTO CONCEPTS & CODE

HTTPS://GITHUB.COM/INSIGHTBUILDER

CHALLENGE SOLVED: DATA REPRESENTATION

WHAT IS ALL THE FUSS ABOUT MATH & MATRICES?

DATA / TEXT / AUDIO/ VIDEO / IMAGE, ALL ARE REPRESENTED AS NUMBERS, INSIDE MATRIX

THIS	IS	NICE
28	57	86
14	82	95
36	07	27
58	68	34
86	26	41

GOOD
2
1
3
2
?

THIS	IS	NICE
28	57	86
14	82	95
36	07	27
58	68	34
86	26	41
7	26	1

PLACE	TO	EAT
7	5	6
48	28	57
67	07	27
85	81	28
76	62	71
?	?	?

HTTPS://GITHUB.COM/INSIGHTBUILDER

CHALLENGE SOLVED: MODEL REPRESENTATION

WHAT IS A MODEL?

$$AX + BY + C = D$$

THIS	IS	NICE
28	57	86
14	82	95
36	07	27
58	68	34
86	26	41

A
B
C

GOOD/BAD/NEUTRAL

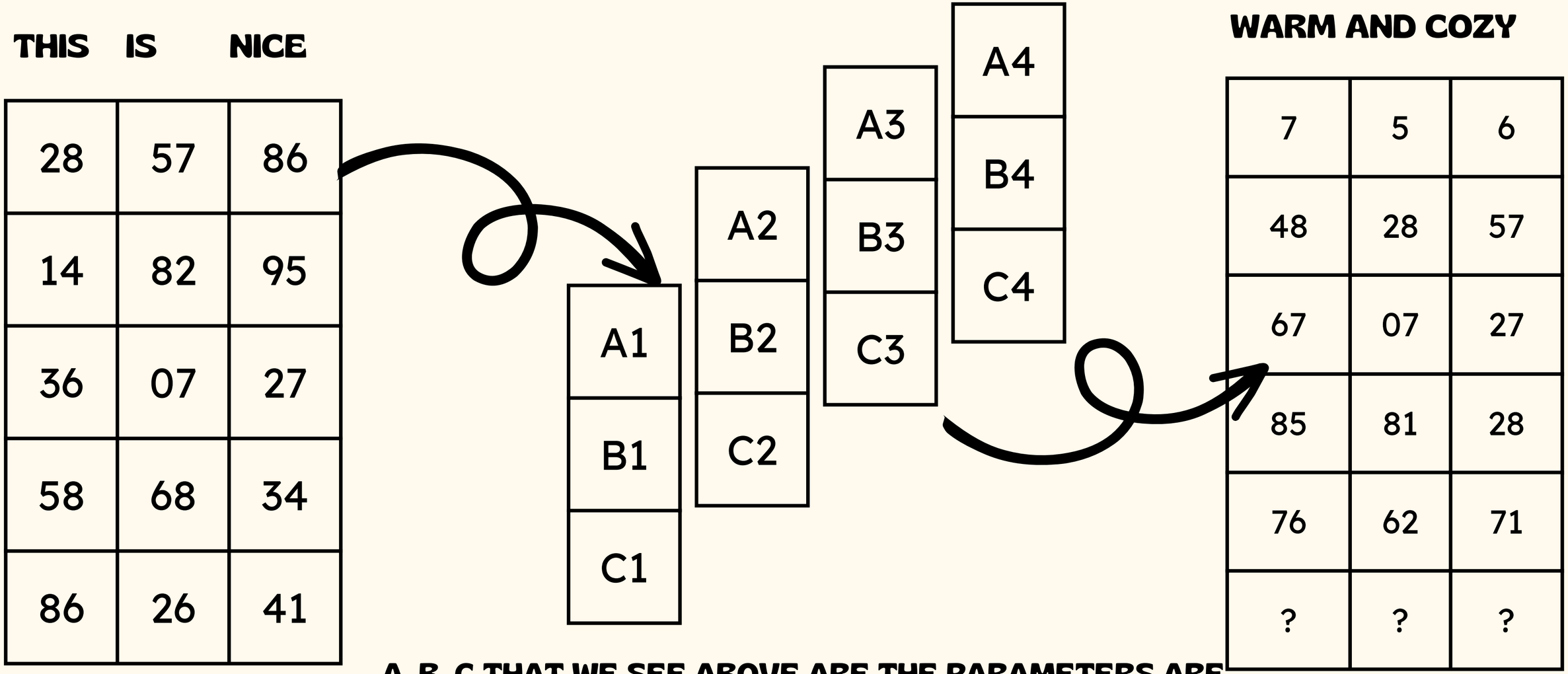
2
1
3
2
?

A, B, C THAT WE SEE ABOVE IS WRITTEN IN FORM OF MATRIX. WHEN THE MODEL IS TRAINED, THESE VALUES ARE RANDOMLY INITIATED AND THE ITERATED UNTIL THE RESULTS ON THE RIGHT CAN BE ACHIEVED

HTTPS://GITHUB.COM/INSIGHTBUILDER

CHALLENGE SOLVED: NEURAL NET REPRESENTATION

NEURAL NET IS A MODEL DESIGN



A, B, C THAT WE SEE ABOVE ARE THE PARAMETERS ARE PART OF THE NEURAL NET. THE TRAINING/FINETUNING PROCESS CHANGES THESE VALUES

HTTPS://GITHUB.COM/INSIGHTBUILDER

CHALLENGE SOLVED: WHAT & HOW OF FINE TUNING

WHAT IS FINE-TUNING?

RE-TRAINING THE NEURAL NET THAT IS ALREADY TRAINED FOR A PARTICULAR TASK

WITH NEW SET OF INPUT DATA & LABELS, SO IT CAN LEARN THE NEW INFORMATION

PROCESS OF TUNING THE NEURAL NET

- DECIDING THE TASK AND THE NEURAL NET THAT NEEDS TO BE TRAINED
- STUDY THE TRANSFORMER IMPLEMENTATION OF THE NEURAL NET
- CREATE DATASET THAT CAN BE USED FOR THE SPECIFIC TASK
- CREATE HUGGING FACE DATASET N PUSH TO HUB (OPTIONAL)
- CREATE TRANSFORMERS TRAINER INSTANCE WITH TRAINING ARGUMENTS
- TOKENIZE THE PROMPTS + EXPECTED OUTPUTS AND FEED TO THE TRAINER
- AWAIT THE TRAINER TO COMPLETE AND THEN
- TEST THE NEW MODEL FOR INFERENCE

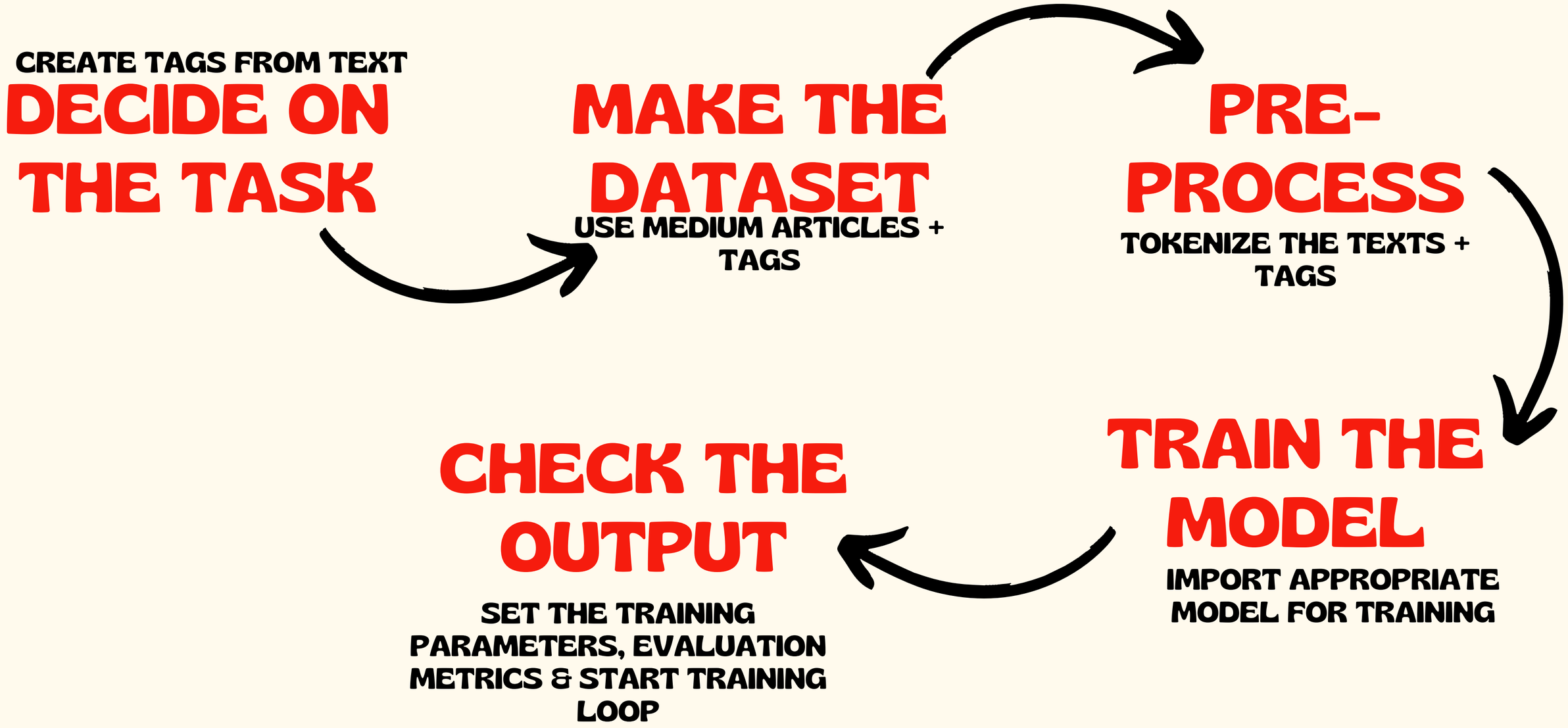
MODEL UNDERSTANDING : 206 MODELS!!!

- EACH MODEL HAS ITS OWN SET OF CONSTRAINTS LIKE MAX_LENGTH OF INPUT/OUTPUT, FORMAT OF INPUT
- TRANSFORMERS DESIGN IS DIFFERENT FOR EACH OF THESE MODELS. EVEN THOUGH EACH MODEL WILL CONTAIN THE ENCODER + DECODER CIRCUIT.
- THESE MODELS ARE EXPLAINED IN DETAIL BY THE RESEARCHERS, AND PROVIDE EXTENSIVE EXAMPLES, AND EXPLANATION ON ITS INNER WORKING AS SCIENTIFIC PAPERS
- HUGGINGFACE TRANSFORMERS LIBRARY HOSTS THESE MODELS, MODEL DETAILS, AND ABSTRACTS ALL THE COMPLEXITY
- HF HOST 206 MODELS. EACH MODEL HAS ITS OWN USECASE, UNIQUE DESIGN FORMULATED BY THE RESEARCHERS AND TRAINED DIFFERENTLY WITH DATASETS
- FLAN-T5 IS IMPROVED VERSION OF T5 RELEASED BY GOOGLE WHICH CAN BE FINE-TUNED & GOOD RESULTS CAN BE EXPECTED
- LEARNING ABOUT THE CORRECT DATA STRUCTURE TO FORMULATE THE DATA FOR FINETUNING
- MIX THE PROCESS OF SUPERVISED & SELF-SUPERVISED TRAINING WITH NOVEL METHODS TO GET THE RESULT THAT IS REQUIRED

REVIEWING THE FLAN T5 NEURAL NET

HTTPS://GITHUB.COM/INSIGHTBUILDER

STEPS TO FINETUNE FLAN T5



HTTPS://GITHUB.COM/INSIGHTBUILDER

HOW TAGS WILL BE CREATED

INPUT TEXT

MANY FINANCIAL INSTITUTIONS STARTED BUILDING CONVERSATIONAL AI, PRIOR TO THE COVID19 PANDEMIC, AS PART OF A DIGITAL TRANSFORMATION INITIATIVE. THESE INITIAL SOLUTIONS
WERE HIGH PROFILE, HIGHLY PERSONALIZED VIRTUAL ASSISTANTS – LIKE THE ERICA CHATBOT
FROM BANK OF AMERICA. AS THE PANDEMIC HIT, THE NEED CHANGED AS CONTACT CENTERS WERE UNDER INCREASED PRESSURES. AS CATHAL MCGLOIN OF SERVISBOT EXPLAINS IN “HOW IT STARTED,

OUTPUT TAGS

"['MENTAL HEALTH', 'HEALTH', 'PSYCHOLOGY', 'SCIENCE', 'NEUROSCIENCE']"

T5-SEQ2SEQLM

MODEL PARAMETERS ARE SHIFTED TO MATCH THE INPUT N OUTPUT

WE DEFINE ACCESS TO A STREAMLIT APP IN A BROWSER TAB AS A SESSION.
FOR EACH BROWSER TAB THAT CONNECTS TO THE STREAMLIT SERVER, A NEW SESSION IS CREATED.
STREAMLIT RERUNS YOUR SCRIPT FROM TOP TO BOTTOM EVERY

LOAD TRAIN MODEL

"[BROWSER, STREAMLIT,SERVER]

HTTPS://GITHUB.COM/INSIGHTBUILDER

THANKS FOR WATCHING

REMEMBER TO PRACTICE WITH EXAMPLES

HTTPS://GITHUB.COM/INSIGHTBUILDER