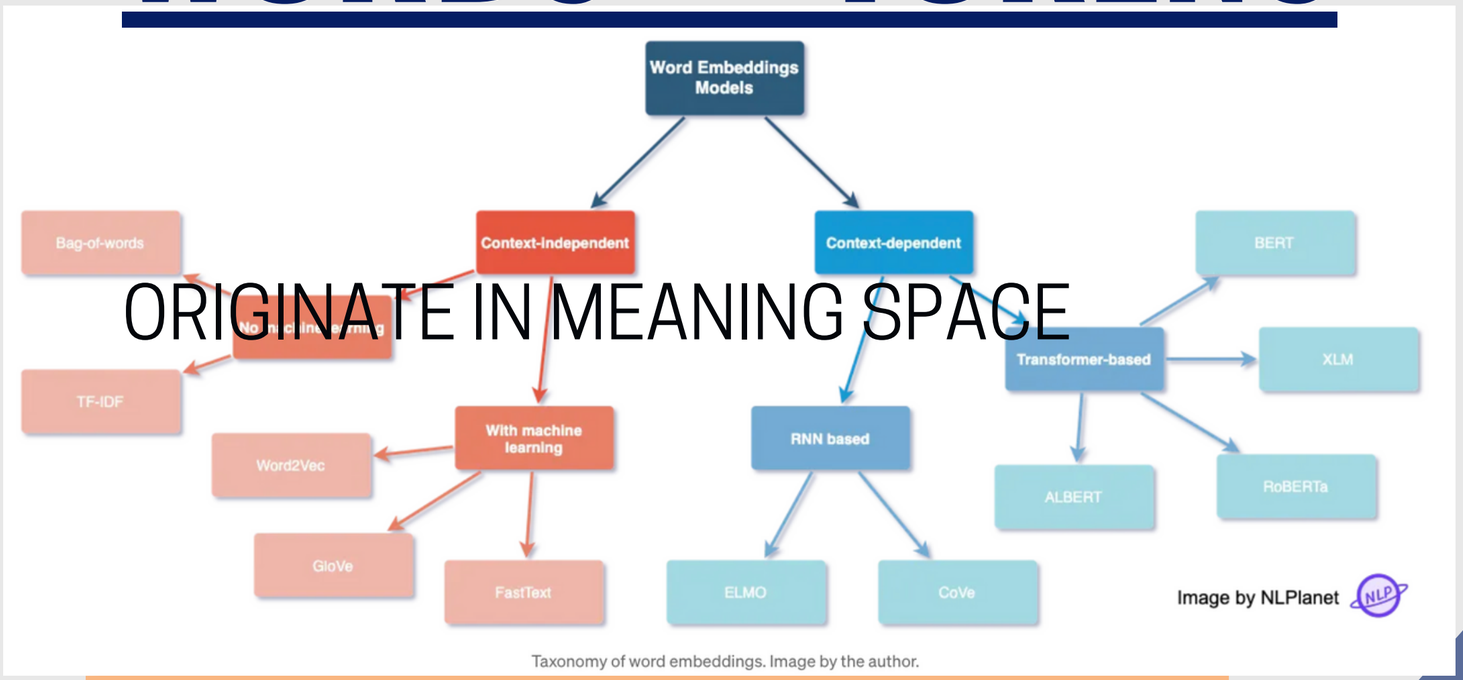
Opinions expressed are solely my own and do not express the views or opinions of my employer.

CS 505
NLP in the Wild

Jena Jordahl,

BU, MS in AI 2023 -> Google 2023

WORDS TOKENS



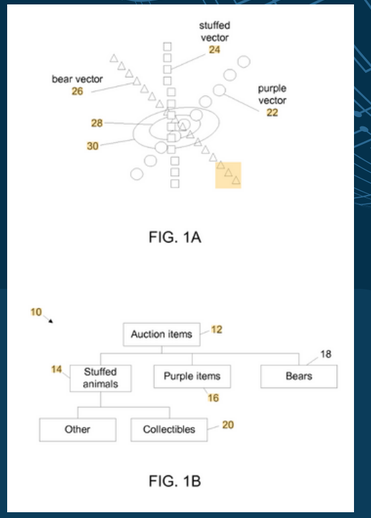
Taxonomy of word embeddings. Image by the author.

ORIGINATE IN MEANING SPACE

Project NLP

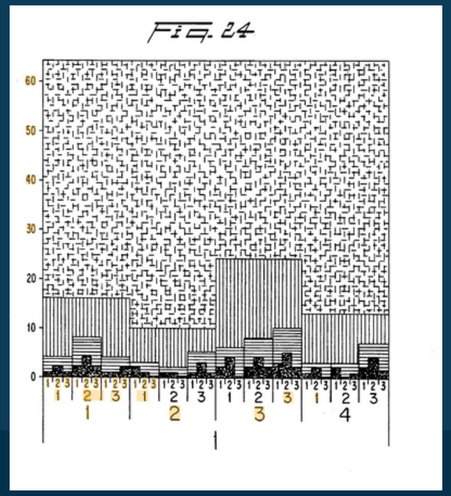
LOOKING FOR MEANING

2001 MULTIPLE HIERARCHICAL POINTS OF VIEW



1990 TEMPLE UNIVERSITY PATENT

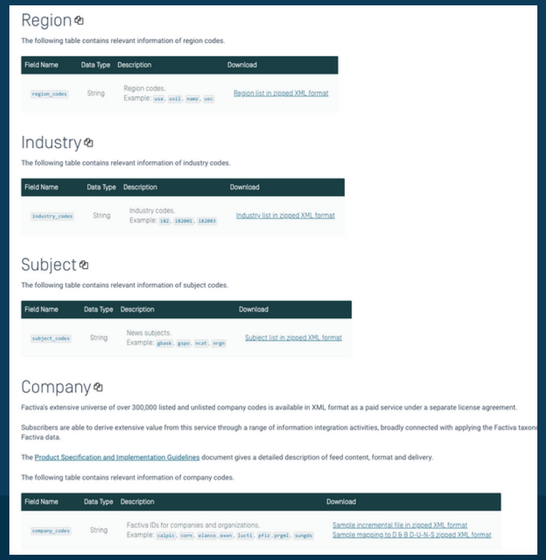
01



DOW JONES CATEGORIES

When trying to solve for 325+ categories

02



Region

The following table contains relevant information of region codes.

Field Name	Data Type	Description	Download
region_codes	String	Region codes. Example: usa, us11, namz, usc	Region list in zipped XML format

Industry

The following table contains relevant information of industry codes.

Field Name	Data Type	Description	Download
industry_codes	String	Industry codes. Example: 182, 182002, 182003	Industry list in zipped XML format

Subject

The following table contains relevant information of subject codes.

Field Name	Data Type	Description	Download
subject_codes	String	News subjects. Example: gbask, gspo, ncat, nrgn	Subject list in zipped XML format

Company

Factiva's extensive universe of over 300,000 listed and unlisted company codes is available in XML format as a paid service under a separate license agreement.

Subscribers are able to derive extensive value from this service through a range of information integration activities, broadly connected with applying the Factiva taxono Factiva data.

The Product Specification and Implementation Guidelines document gives a detailed description of feed content, format and delivery.

The following table contains relevant information of company codes.

Field Name	Data Type	Description	Download
company_codes	String	Factiva IDs for companies and organizations. Example: calplc, corn, alance, exxn, luct1, pfiz, prgel, sungdc	Sample incremental file in zipped XML format Sample mapping to D & B D-U-N-S zipped XML format

MIKOLOV 2013 WORD2VEC

03

Czech + currency	Vietnam + capital
koruna	Hanoi
Check crown	Ho Chi Minh City
Polish zolty	Viet Nam
CTK	Vietnamese

ELM0 CONTEXT DEPENDENT 2018

04

Pre-trained word representations (Mikolov et al., 2013; Pennington et al., 2014) are a key component in many neural language understanding models. However, learning high quality representations can be challenging. They should ideally model both (1) complex characteristics of word use (e.g., syntax and semantics), and (2) how these uses vary across linguistic contexts (i.e., to model polysemy). In this paper, we introduce a new type of *deep contextualized* word representation that directly addresses both challenges, can be easily integrated into existing models, and significantly improves the state of the art in every considered case across a range of challenging language understanding problems.

Project MathQ

IDDO DRORI AI COURSE

A Neural Network Solves, Explains, and Generates University Math Problems by Program Synthesis and Few-Shot Learning at Human Level

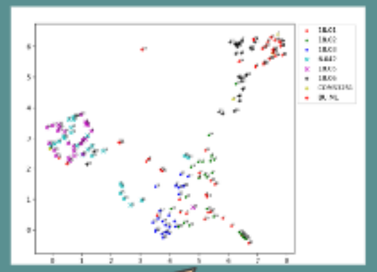
OOPS!

Did not receive permission to release the data or model fine-tuned on the data

We evaluate the ability of large language models to fulfill the graduation requirements for any MIT major in Mathematics and EECS. Our results demonstrate that GPT-3.5 successfully solves a third of the entire MIT curriculum, while GPT-4, with prompt engineering, achieves a perfect solve rate on a test set excluding questions based on images.

Autocast Competition

OpenAI, embedding
sentence_transformers



MIT Courses

BU.ML & BU.AI

1. Verified mathQ

Reproduced paper results and added new courses. This process taught us methods to creatively interact with ChatGPT to get the best predictions.

question_embeddings.json

Train

Test

Remove Test data from Training Data

Training

Autocast

Created a directory, one question per file

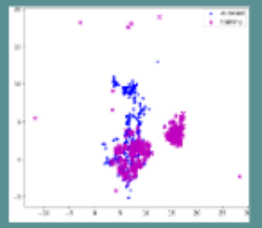
The FutureSight GAME
2) Adapt Zero-shot and Few-shot Learning ->

3) Multiple response cycles to format, capture and evaluate the answers

4) Two cycles generating and applying summary and argument ChatGPT responses the specific to each question

Game Play learning Method

5) Setup the system prompt to give the context of the game to enforce format of the answers and willingness to give predictions.



6) Using response cycles to extract the answer from text and output to a spreadsheet

Separate the autocast questions based on the date of the model's training data. Cutoff Midnight the day before the close_time.

GPT-4 trained Sept 2021 =794

GPT-3 text-davinci-003 = 347

Davinci trained Oct 2019 = 223

7) Limited capability

ABOUT GOOGLE CLOUD'S JOB

Our team of Generative AI Blackbelt experts is dedicated to the success of Google's elite 200 client portfolio. We are skilled at unlocking a trove of productivity enhancements through innovative LLM technical solutions. We lead engaging Executive Briefings and immersive tutorial sessions, where we shine a spotlight on the transformative potentials of Generative AI tools for language and image analysis. We reveal the power of innovation to redefine the productivity landscape.

About Kubra Eryilmaz

MICROSOFT

Leverages transformative possibilities for global clients by providing innovative solutions that seamlessly integrate LLM APIs into their unique business frameworks. Harnesses the power of technology to boost efficiencies, streamline operations and realize your enterprise's potential. Holds an engineering position in the middle of a sales organization.

Learn More

Table Of Content

01. Embedding Spaces
When meaning makes High Dimensional Spaces Easier

02. MathQ Prompt Engineering and Beyond
Sophisticated Prompts and Multi-Turn approaches with Auto-Evaluation

03. Tale of two BU Graduates & LLMs
The path from your past and BU's projects to your work life

04. Chatbots & RAG Retreival Augmented Generation
Combining Search and LLMs

05. Multi-Modal LLM
Car Damage Detection AI Model

LLM - Dialog Chatbots

Customer service could not get much worse so maybe it will get better with the new round of chatbots. This is likely the most prevelant use of LLMs in the near future.

Blender
Custom face and features creation

Hair, Clothing and Makeup
Multiple options to choose from

Voice
Multiple NLP options supported

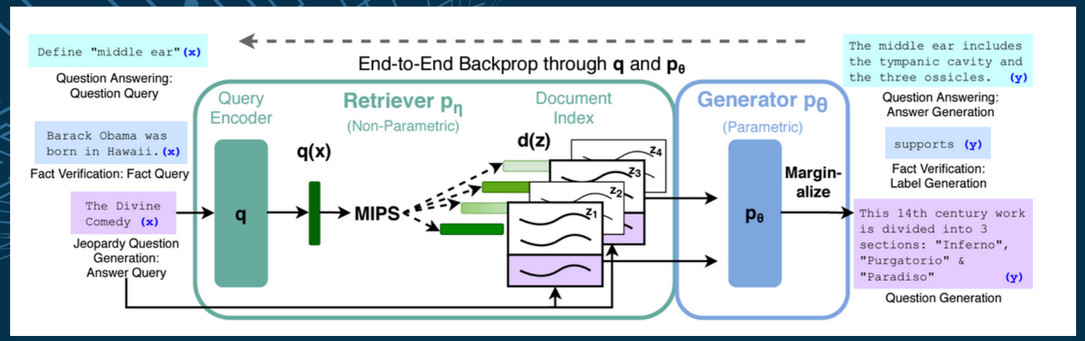
Personality
Choose, configure and refine

Gestures
Personalize with multiple gesture options

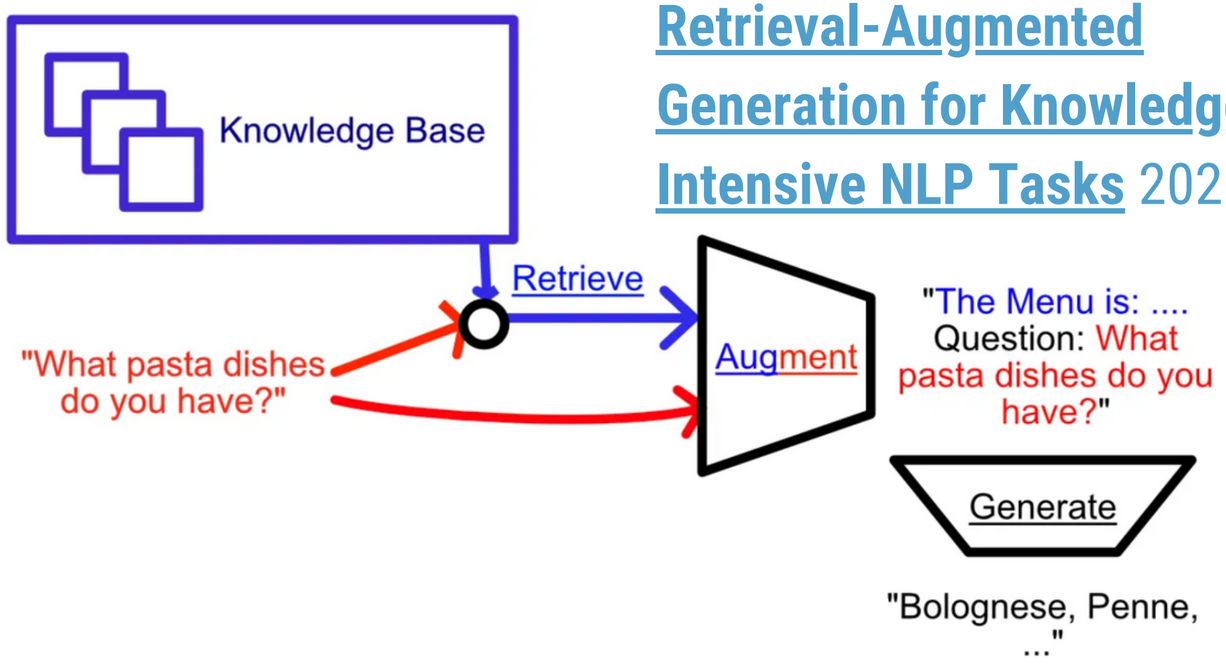
Conversation
Create your own using Google DialogFlow, IBM Watson, Microsoft Azure Bot Service, and Amazon LEX or other NLPs. Or leverage our Open AI GTP integration, or a combination.

Deployment
Deploy across multiple digital platforms and screens or export to video for social sharing.

The only really conceptually challenging part of RAG is retrieval:
How do we know which documents are relevant to a given prompt?

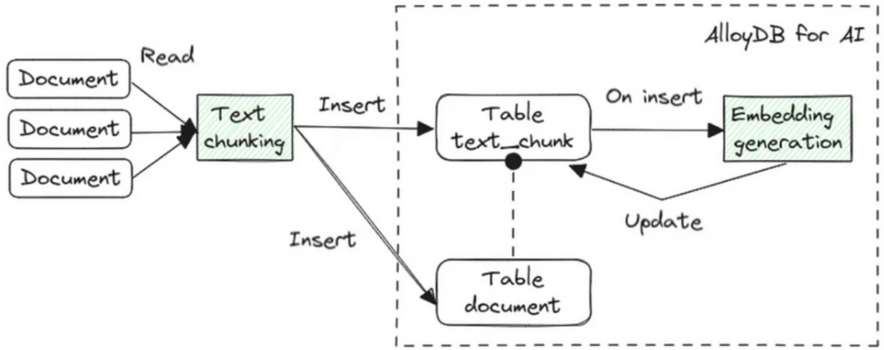


RAG = LLM - Search



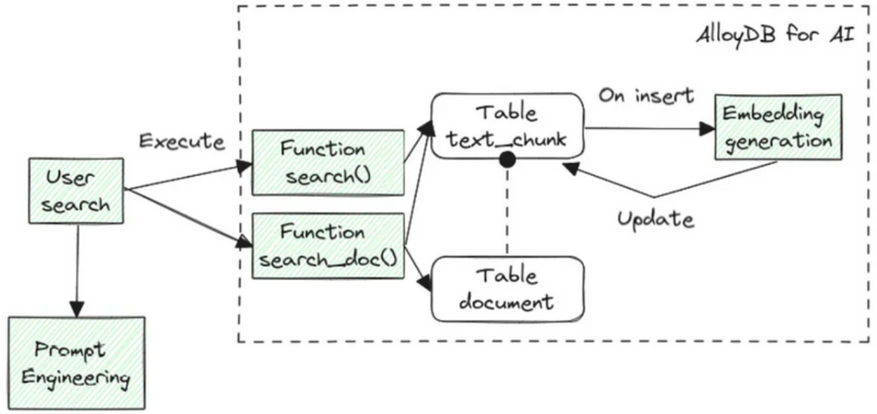
Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks 2021

The workflow for embedding generation and storage is (the dashed line is a foreign key relationship):



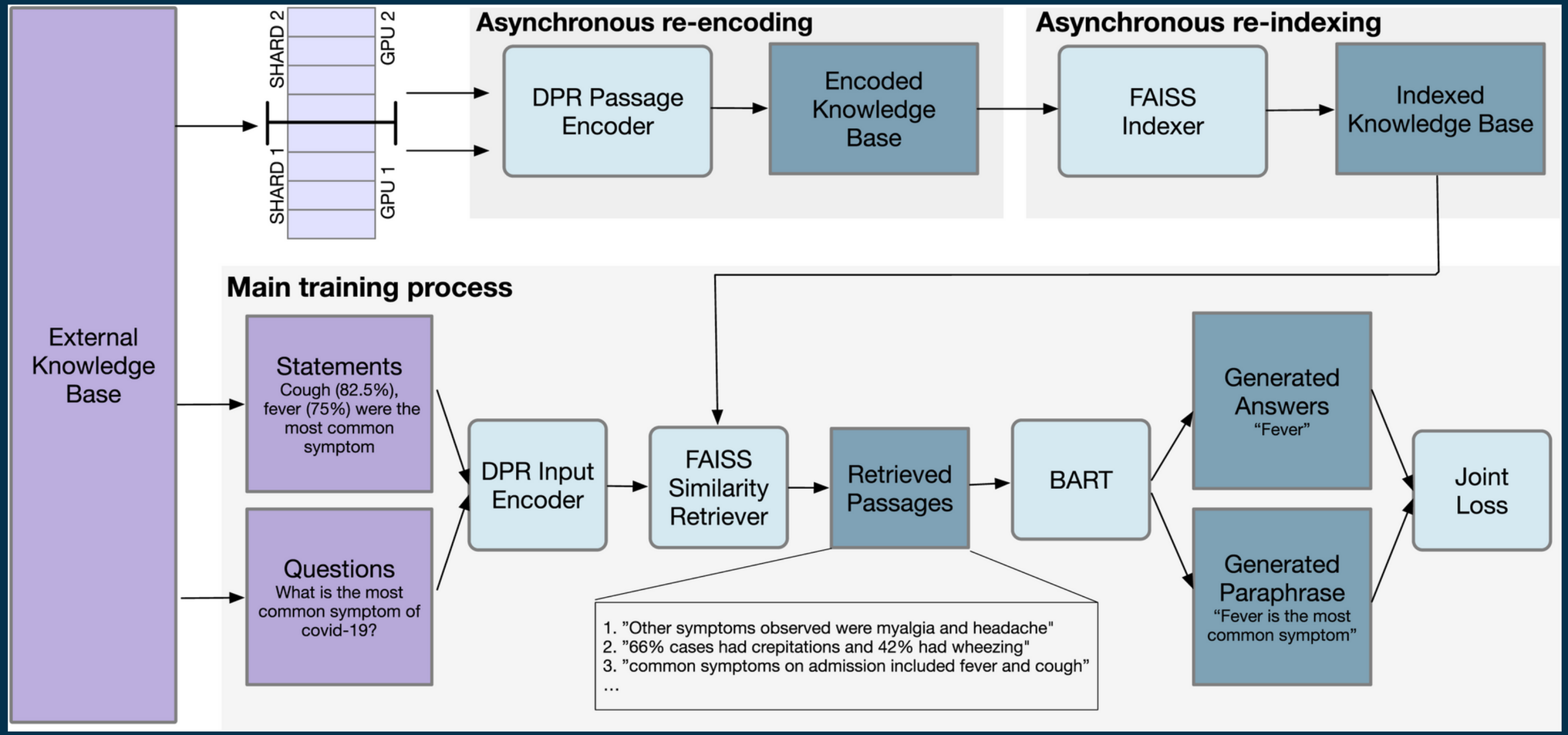
Embedding Generation and Storage

The workflow for query execution is:



Query Execution ("search")

Transactions of the Association for Computational Linguistics. 2023;11:1-17. doi:10.1162/tacl_a_00530



System Overview. Our RAG-end2end training architecture uses asynchronous processes to dynamically re-encode and re-index the knowledge base while optimizing a joint QA and paraphrasing signal loss. The training dataset consists of both reconstruction signals and QA pairs. The network learns to generate answers to questions and useful statements jointly. The input to the BART reader is illustrated in Equation 3, where the model can differentiate the answer generation task and statement reconstruction task with the use of a control token. During the training, embeddings and the knowledge base index get updated asynchronously.

Our ReAct: Synergizing Reasoning and Acting in Language Models

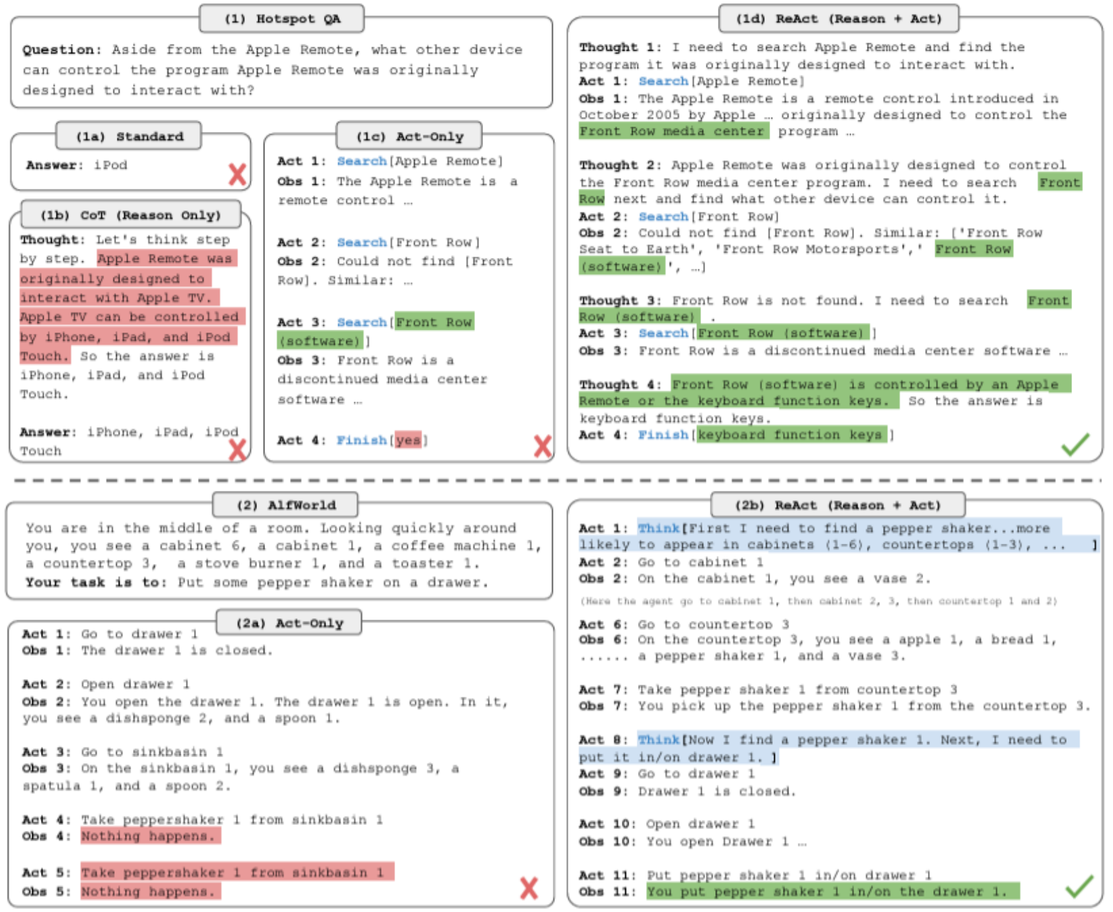


Figure 1: (1) Comparison of 4 prompting methods, (a) `Standard`, (b) Chain-of-thought (`CoT`, Reason Only), (c) `Act`-only, and (d) `ReAct` (Reason+Act), solving a HotpotQA (Yang et al., 2018) question. (2) Comparison of (a) `Act`-only and (b) `ReAct` prompting to solve an AlfWorld (Shridhar et al., 2020b) game. In both domains, we omit in-context examples in the prompt, and only show task solving trajectories generated by the model (Act, Thought) and the environment (Obs).

ML Beginner's Guide to Build Car Damage Detection AI Model

Multi-Modal

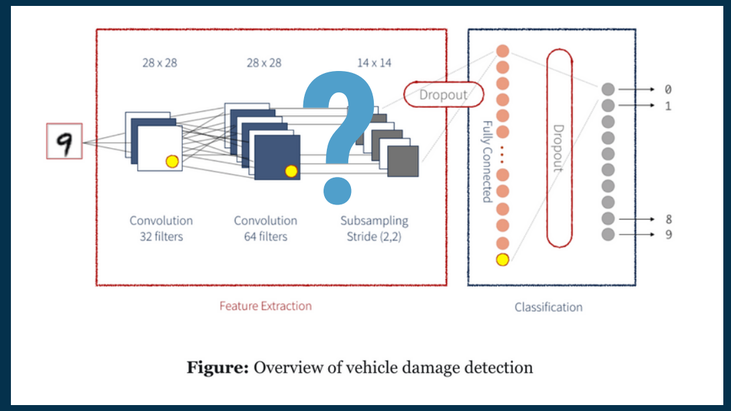
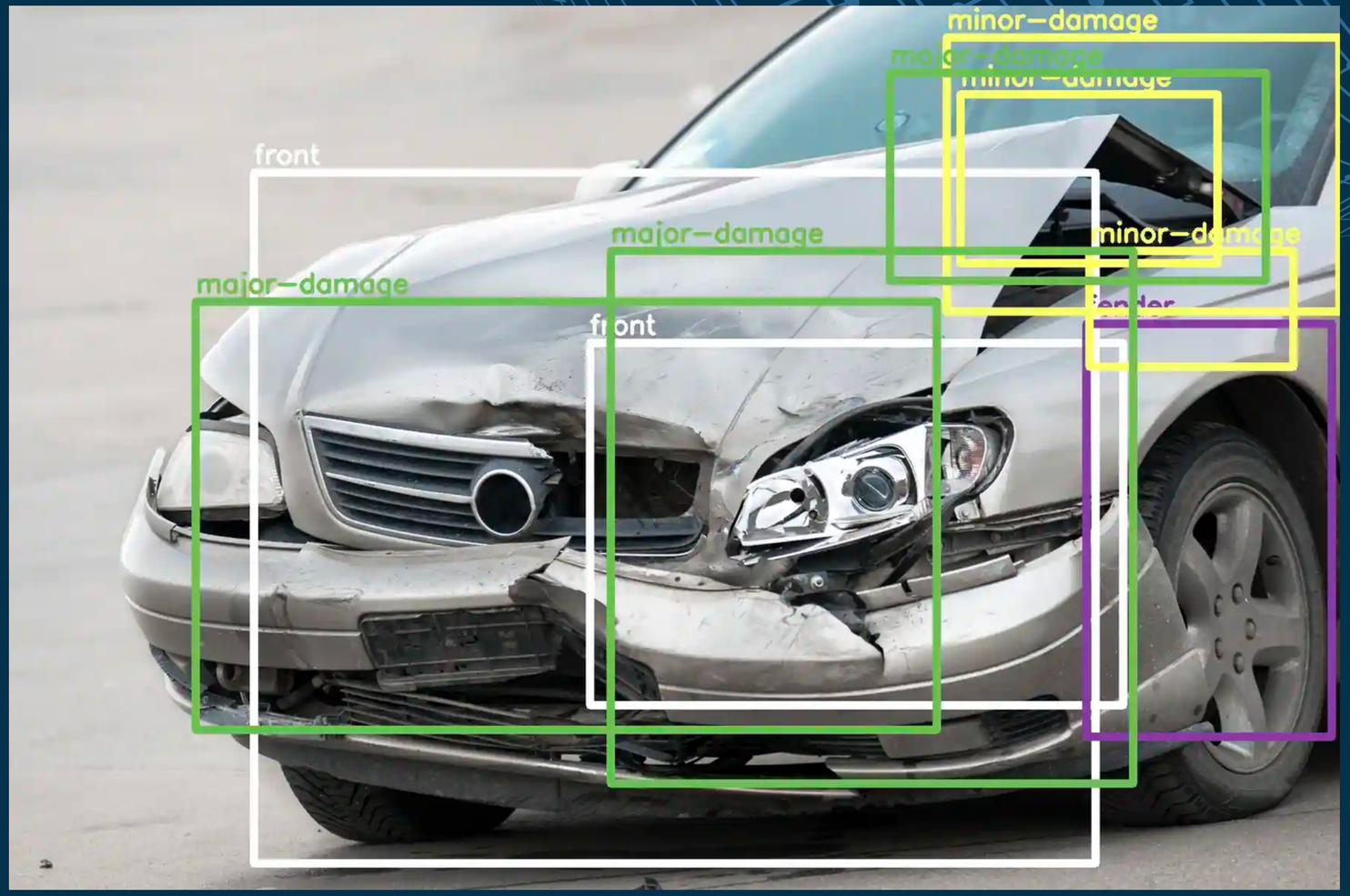


Figure: Overview of vehicle damage detection



Thank You

jjordahl@google.com