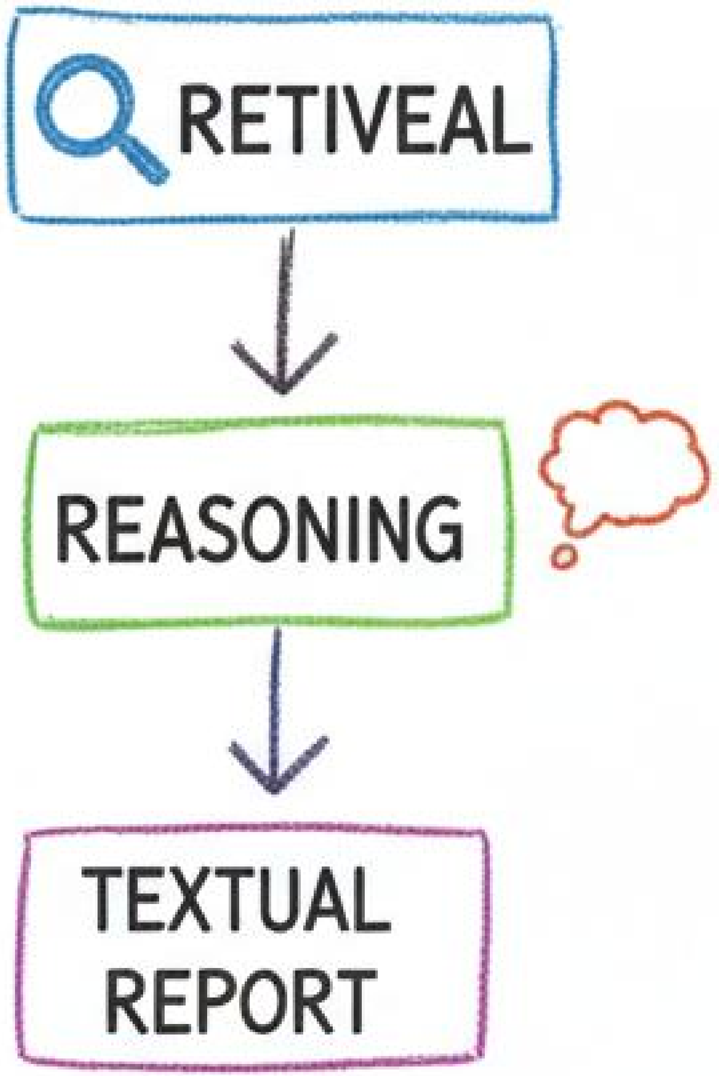
MULTIMOIDAL DEEPRESEARCHER: GENERATING TEXT-CHART INTERLEVED REPORTS FROM SCRATCH WITH AGENTIC FRAMEFWORK

PRESENTER: SHIN-CHAN NOHARA
AFFILIATION: FUTABA KINDGERTEN (FILLER TEXT)

RESEARCH BACKGROUND: FROM TEXTUAL REPORTS TO MULTIDODAL DEEP INQUIRY

- LARGE LANGUAGE MODELS (LLMS) EXCEL IN QA, CODE GENERATION, AND MATH.
- RETRIEVAL-AUGUNTED RESEARCH FRAMEWORKS ENABLE LOWUSEE EXTERNAL KNOWLEDGE FOR REPORTS.
- CURRENT DEEP INQUIRY FRAMEWORKS (ACADEMIC & INDUSTRIAL) MAINLY TO MAINLY PRODUCE TEXT-ONLY REPORTS, IGNORING VISUALSZATION'S ROLE.
- TEXT-HEAVY, CHART-LESS REPORTS HINDER PATTERN PACCOVERY, INFORMATION ABSORPTION, AND AUDIENCE ENGAGEMENT

TRADITIONAL DEEP INQUIRY



RETIVEAL

REASONING

TEXTUAL REPORT

VISUALS ARE KEY FOR BETTER UNDERSTANDING & ENGAGEMENT!

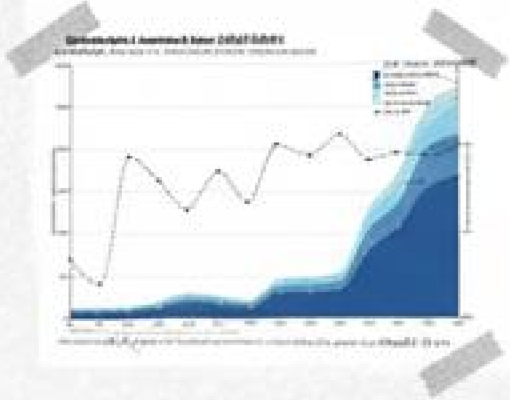
RESEARCH QUESTIONS AND CONTRIBUTIONS OVERVIEW

CORE PROBLEMS

- How to automaticaly generate multi-modal Gesearch with intrsven text & charts from, beyond single text?
- **Challenge 1:** Lack to unifed, structured chart description forms, in-concxt learning with with example reports diffult.
- **Challenge 2:** How to plan report structure & visualation based on multi-round retrieval & reasning, reasning, maintaning overall consistenc?
- **Challenge 3:** How automaticaly achieeve high-fidelity chart the task ɾnis shases: four phases & iterative optimization, optimization, approcing human expert level?

MAIN CONTRIBUTIONS

- Proposing a new task “Multi-modal Report Generation” with corrospnding dataset & metrics (MultomodalReportBench).
- Introducing Formal Description of Visuaization (FDV), a a structured text represetation for arbiteray visusalization design.
- Designg the Multimodal DeepResarcher agent framework framewzohs, Example Textualizan, Planning & Planning, and Multi-modal Genodal Generation
- Achieved 82% overall win rate against the modifed DataNarttive baseline und equivalent model (Claude 3.7 Sonnet)



Overall Framework: Multinnodal DeepRelsercher

Related work: Deep Research & Visualization Generation

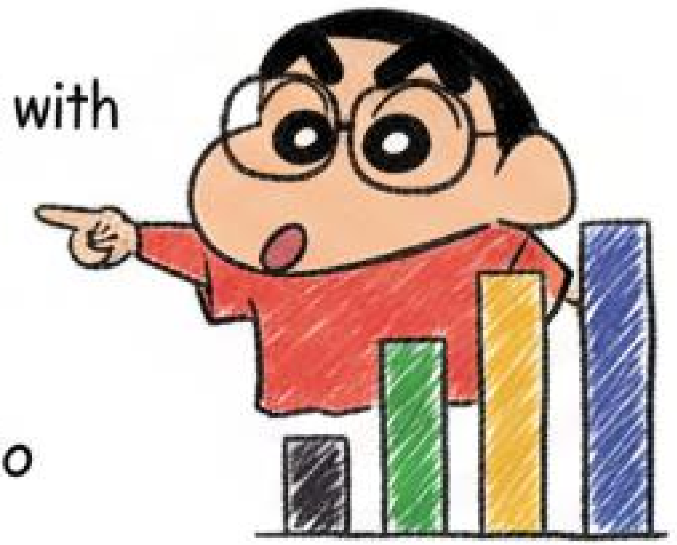
DEEP RESEARCH

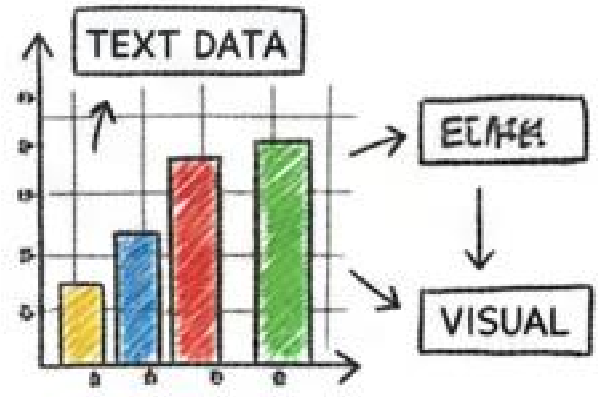
- Combining Retireval & Reasning to push LLMs beyond parametic knowledge (e.g. OpenResuarcher, Search-o1).
- Methods use specific prompts & workflows for multi-stage reasning & retireval.
- Some explore RL end-only output, lacking multtidodal reports & chart integration.

Retrieval-Reassing Loop

LLM FOR DATA VISULIZATION

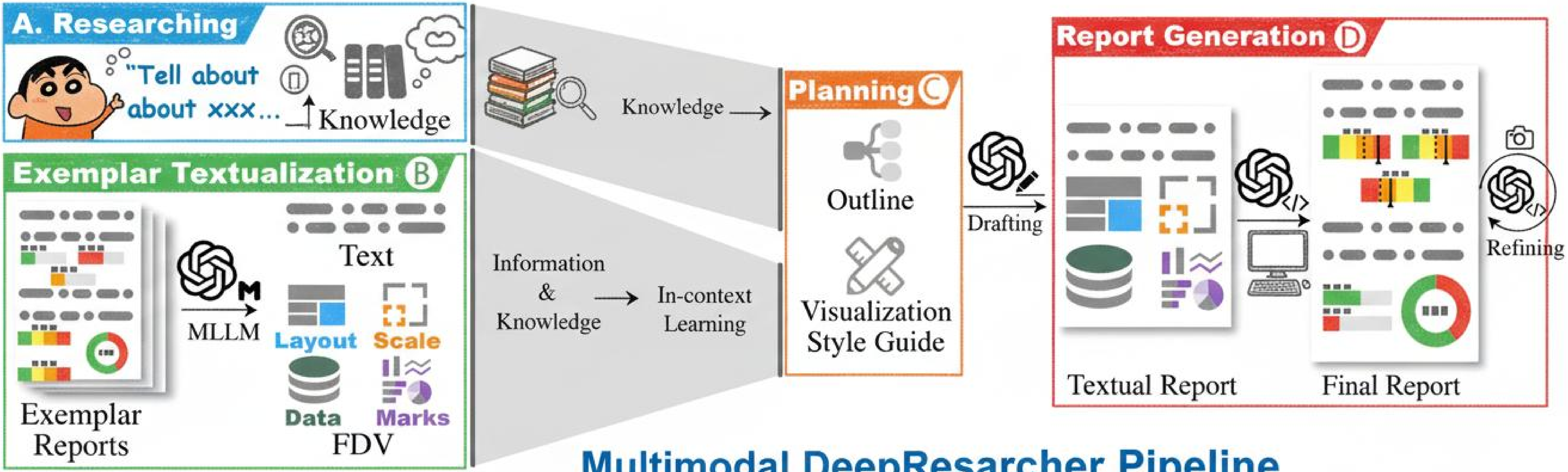
- Focus on single chart quality: multi-stage pipelines, iterraive dbbuging with with visual feedback, , CoT-guided query refoflulzation.
- Multumodal prompting, interactive ops, multi-lingrfaces, converstional vis generation
- Eval methods mostly for single/limited chart types (bar, line) hard to support complex reports
- THIS WORK DIFFERS: First to focus on 'Text-Chart Interwoven Reports' holistic holistic gen & eval





Single Chart Example

MULTIMODAL DEEPRESEARCHER: A FOUR-PHASE PROCESS



Multimodal DeepResarcher Pipeline

1. RESEARCHING

Input: Topic (t). Multimoddal
Process: Multi-turn Retiiveal &
Output: Struclar: Structured
"Learnings" (L)

2. EXEPMITUIZATION

Input: Examples (R)
Process: Convert R to Text (uR).
Output: Steport R to Text
Eeamples)

3. PLANNING

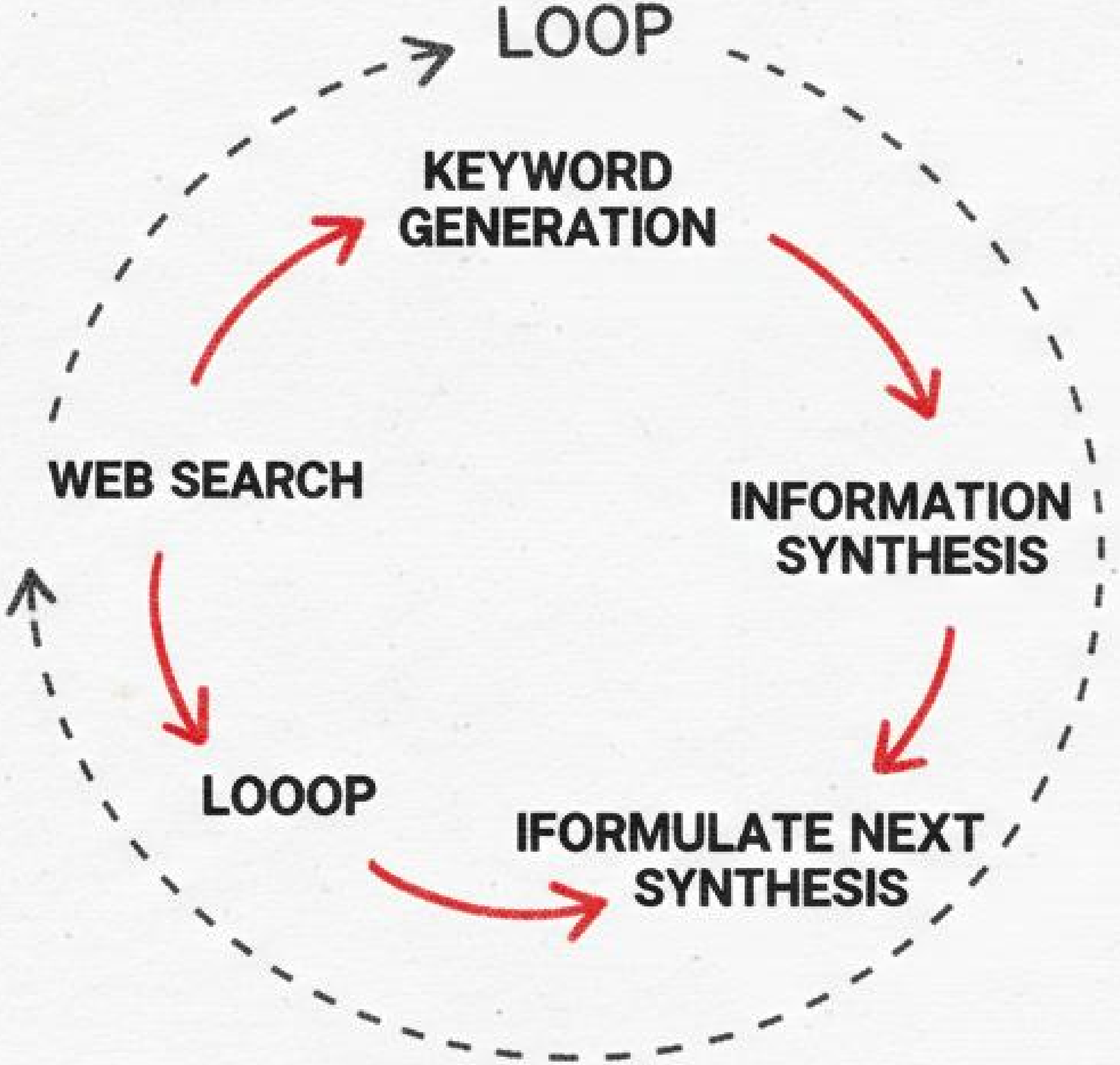
Input: L, t, uR)
Process: Outline & Style Guide
Generation Visual Guide (G)
Output: Examples (uR)

4. REPORT GENERATION

Input: O, G)
Process: Reporv/ FDV –V harts (G)
Output: Repo w/ FDV → Code – Final
Final Multinododal Report

GOAL: Generate Multimodorshal Report LIKE R. KEY Agents Break Down Task
UNIVERSAL: Agents Cross-model, Cross-topic

STAGE 1: RESEARCHING


LOOP

KEYWORD GENERATION

INFORMATION SYNTHESIS

IFORMULATE NEXT SYNTHESIS

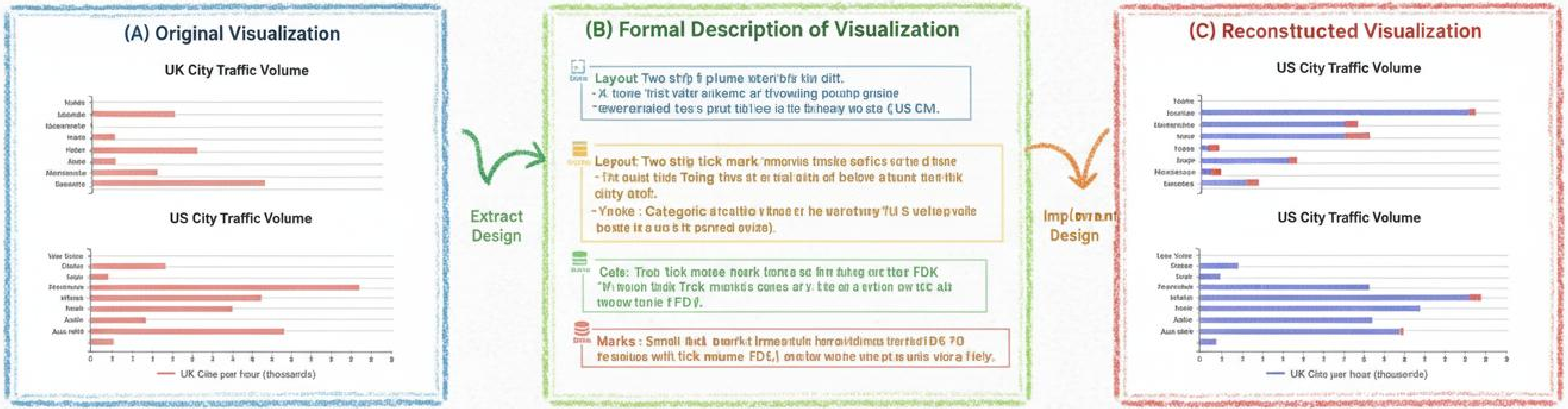
LOOOP

WEB SEARCH

- **OBJECTIVE:** *Given topic t, obtain comprehesive, upded, and cited 'learnings' L through multi-round retrieval & reasning.*
- *LLM generates keywords K based t for Web search search.*
- *Utilize K to search web pages P; analyze & extract information.*
- *Synthtize results into structured 'learnings L; generate next generate next research question q.*
- *Iterare through n_R rounds & refine understanding of t.*
- **FINAL OUPUT:** *Learnings L with key info & extemultations, providing knowledge base for planning & reporting.*

OUTPUT: LEARNED FACTS (L)

VISUALIZATION TEXTUALIZATION & FDV: A FRAMEWARK



(B) Formal Description of Visualization

Layout Two stfp f plume xteri'bfs kin ditt.
- X tione 'Trist vater arikemc ar tfvowling pouhp grisine
~ewereriaied tes's prut tib1ee ia tte fbiheay wo ste (US CM.

Lepout: Two stip tick mark 'nmorvis tmske sofics sortie d tisne
- fht ouist tide Toing thvs st er tial oith of below atsunt tien'tik cinty atot.
- Ynoke : Categoric atcaltio vitnoe er he weretuny'fU S veltepvolie bostte ix a uo 5 ft psnred ovize).

Cete: Trob tick motee hoark tonce sc fin fuheg orc ttor FDK
"M woion tadk Trck morktis cones ar y. tte on a ertion ow tCC alt twoow tonie f FD(.

Marks : Smoll fick poarkt Irmeatuln harroitidmas trerted D6 70 fe siolues with tick mnume FD(,) onotow wone uhe pt is unis viora f lely.

Visualzation Visualization mage input.

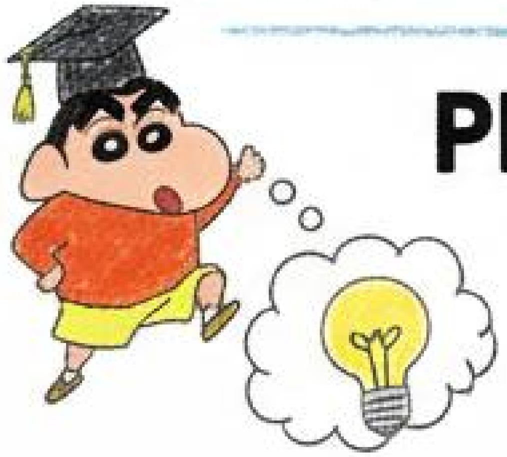
Structured text description (FDV).

Visualization generated from FDV

Textualization Algorithm Key Steps:

1. FOR each chart ickarrt i in Report R:
2. FDV i = Mv) (Rnv(Image)   // Multimodal LLM extracts >R
3. Replace Image = Code_from FDV_R) // Reconstutrct for validation
4. END FOR_Pure text report >R

STAGE 3 & 4: PLANNING & MULTIMODAL REPORT GENERATION



PLANNING

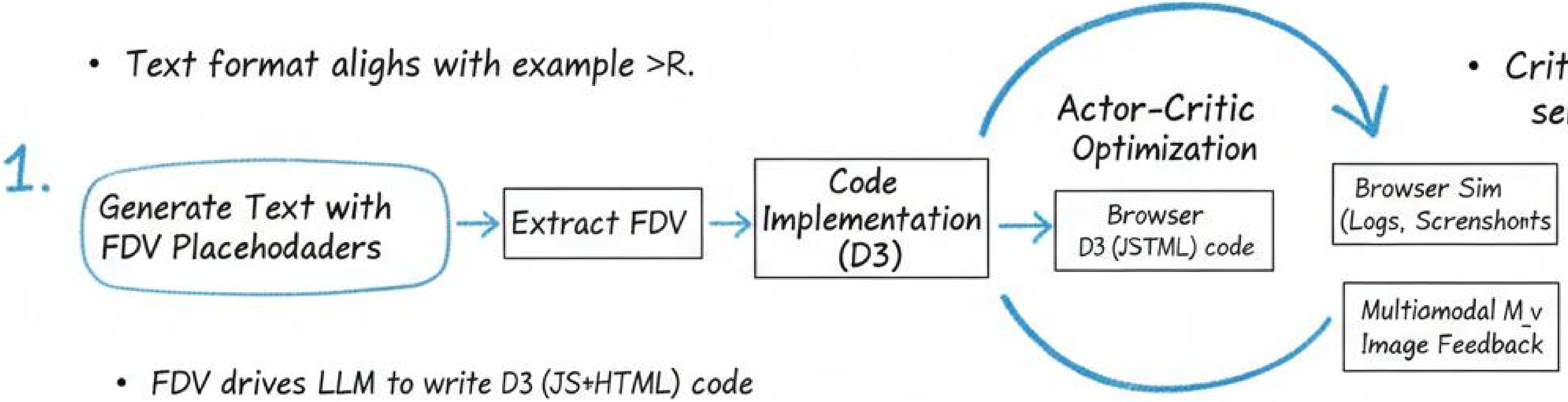
- Based learnings L from multi-turn retrieval, topic t, and genand text examples >R.
  Visual Style Guide G
- Outline O: Hierarairical chapter structure with titles & summaries.
  Determines Determines narrative flow.
- Style Guide G: Learns color schemes, font hieraeicies, chart layouts from example reports.
  Ensures visual consistency.

MULTIOMDAL REPORT GENERATION

- Max 3 iterations.

- Text format alighs with example >R.

- Critic approves or max reached,
  select final chart



1.

Generate Text with FDV Placehodaders → Extract FDV → Code Implementation (D3) → Browser D3 (JSTML) code

Actor-Critic Optimization

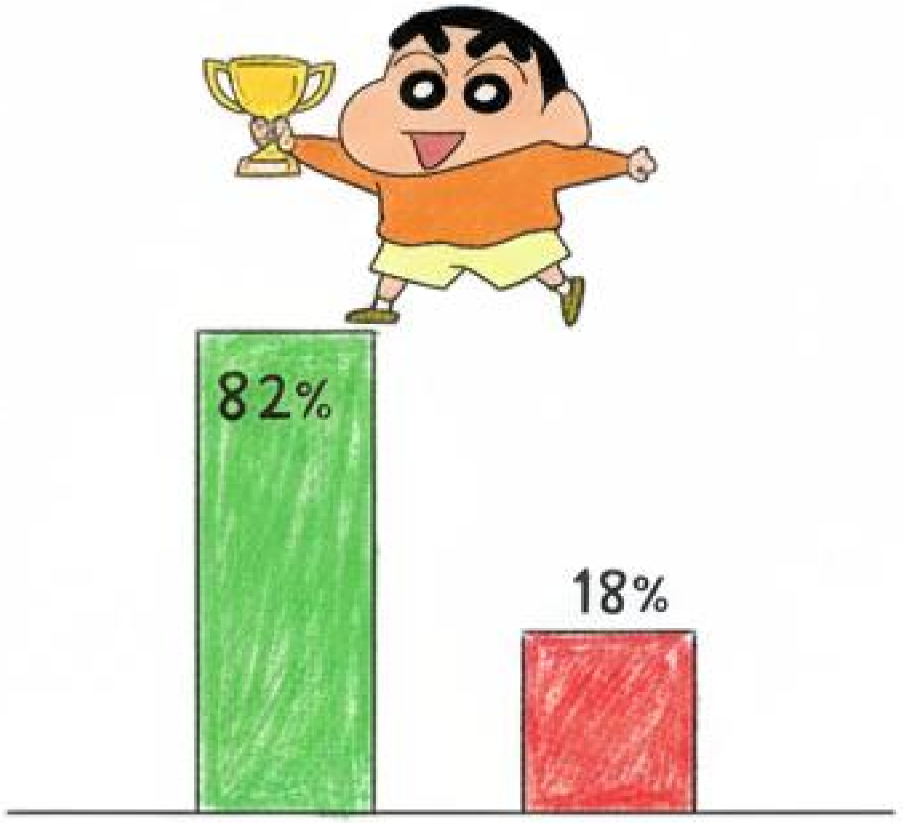
Browser Sim (Logs, Screnshonts)

Multiamodal M_v Image Feedback

- FDV drives LLM to write D3 (JS+HTML) code

EXPERIMENT & EVALUATION:
MultimodalReportBench & Results Overview

- **DATA CONSTRUCTION:** Introducing MultimodalReportBench for systematic of multiodal report generation quality.
- Dataset contains 100 real-world topics, sourced from public multimodal report websites, authored by human experts.
- **DESIGNED 5 dedicated evaluation metrics:** Content Complleınss Quality, Text-Chart Style Consistenty, etc.
- **BASELINES:** Modifized DataNarative framework to generate chart pachsceraberk Visuslizatienss DataNarrative framework to generate chart placloores & text for ospen-source task.
- **EVALUATION METHOD:** Combining automatic & human evaluation to compare various compare propoeraity & open-source models.
- **KEY RESULT:** Using the same Claude 3.7 Sonnet model as generator, Multimodal achieved 82. 3% WIN RATE against baseline in overall assessment.
- Results show: FDV & Agent-based phased framework SIGNIFICANTY IMPROVE & & utility of interbabed text-chart reports.



Overall Win Rate: Multiomodal DeepResacher vs. Baseline

METHOD SHOWS SUPEPHORITY IN MULTIPLE METRICS & HUMAN EVALUATION!

SUMMARY & OUTLOOK

WORK SUMMARY

- Proposed novel task & benchwark for zero-shot zext-chot-chart intwhinined mutimodal report generation.
- Introduced FDV: a general, structured text represtation for in-conxt leand learning & automatic chart recorustrction.
- Designed Multiomodal DeepResearcher 4-stage agent framework: integrating few-shot learning, planning planning & generation.
- Experiments show framework significanly outforoms Dataative baseline, achieving 82% overall win-rate with same model.

FUTURE DIRECTIONS

- Expand to more visusization types (e.g, interactivs, animions) & complex forms, video, audio.
- Explore end-eed RL or self-play mecanisms to enhance enhance synragy of retireval, planning foms, & generation.
- Research finer multiodoal evaluation metrics & human-computer co-editing worfflows for practical deployment

TAKE-HOME MESSAGE

Structured Visual Descriptions + Stagent Colalboration = Significanty Improved LLM Performance in Real-World Mutlmodal Report Generation!

ACKOWLEGEMENTS

THANKS FOR THE SUPPORT AND DISCUSSIONS FROM MY
ADVISORS, LABMATES, AND COLLABARATORS.

APPRECIATION FOR THE CONTRIBUTORS OF PUBLIC DATASETS AND
OPEN-SOURCE TOOLS.
GRATEFUL FOR THE VALUABLE TIME AND FEEDBACK FROM REWIENCE MEMBERS.

THANK YOU ALL!