NLP Project

ENRICHING LLMS WITH RAG

Jan Henrik Bertrand, André Gilbert, Felix Noll, Marc Grün

Try Pitch

01

Motivation

02

Explanation

03

Demonstration

04

Evaluation

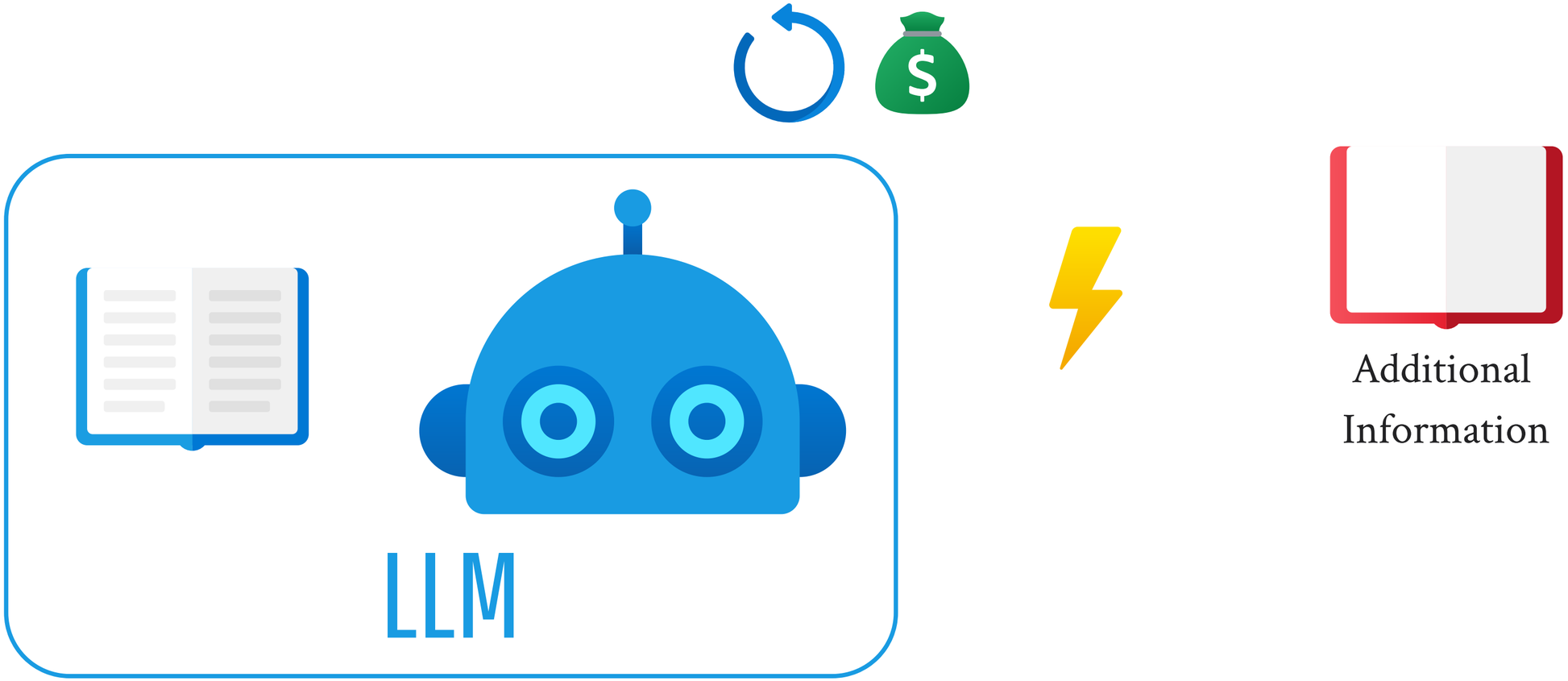
Try Pitch

Section 01

MOTIVATION

Try Pitch

WHY USE RAG?



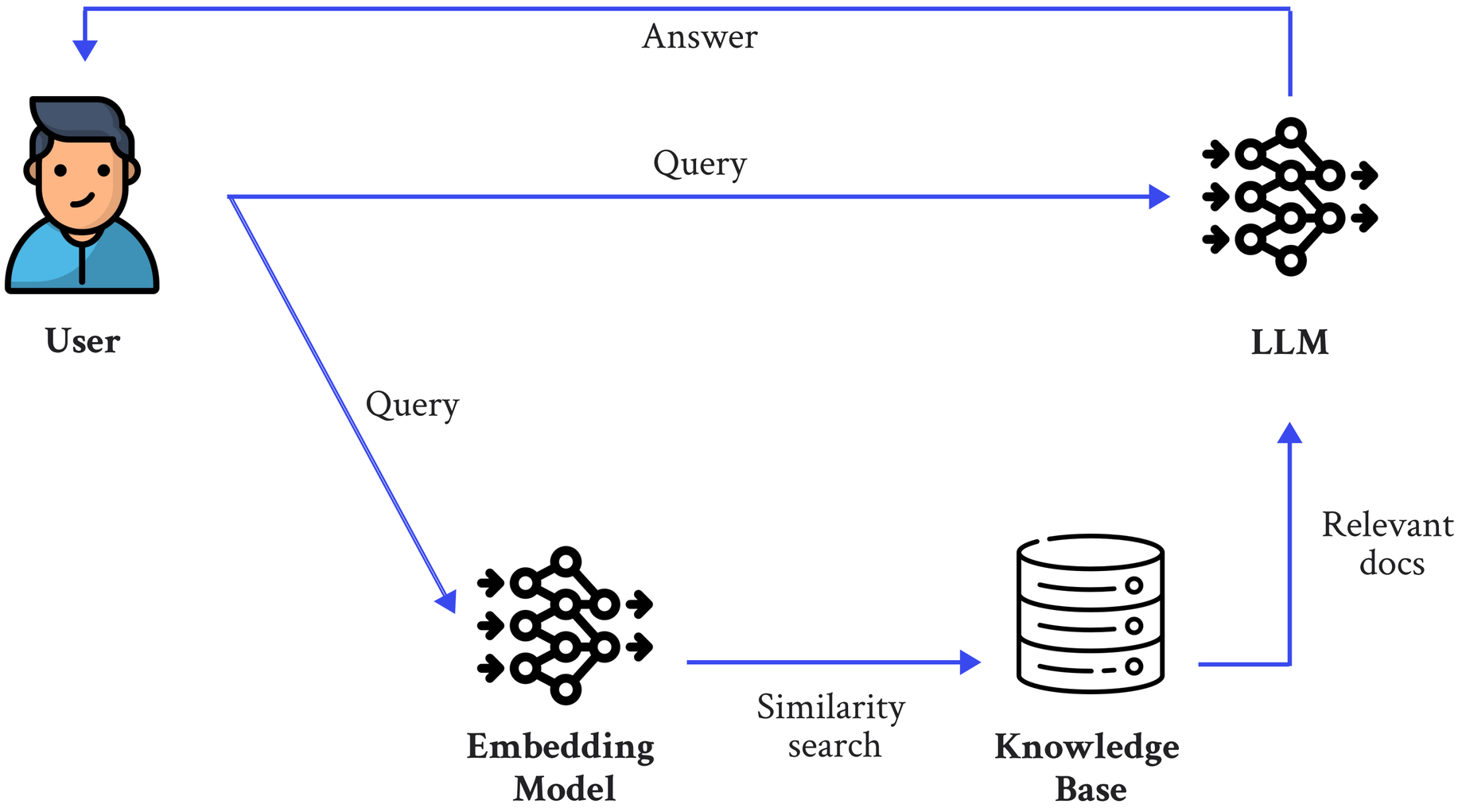
Try Pitch

Section 02

EXPLANATION

Try Pitch

RETRIEVAL AUGMENTED GENERATION

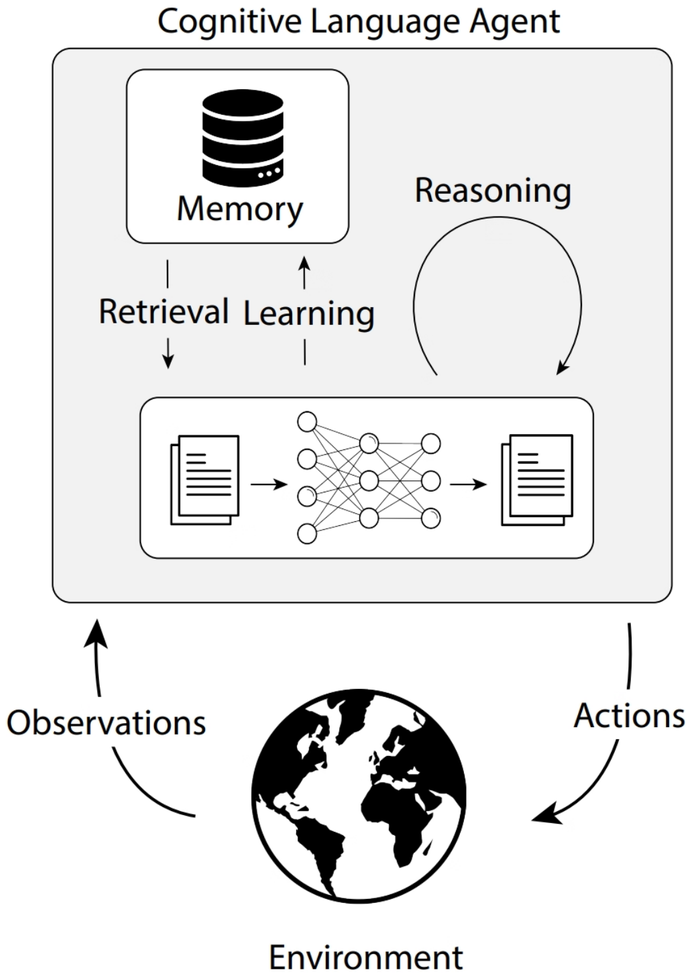


Try Pitch

REACT PROMPTING

ReAct (Reason + Act) Prompting

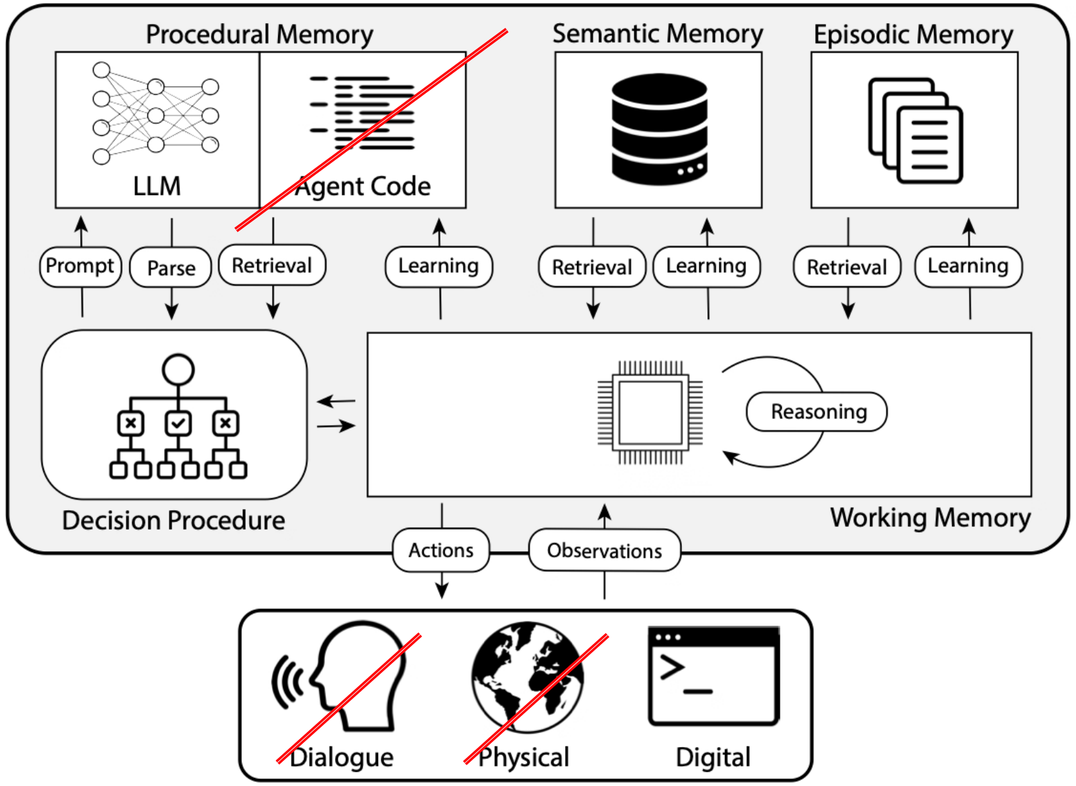
User Query: Read in the test.csv as a pandas DataFrame.
Thought 1: I don't know pandas but since the required output is a pandas DataFrame, I will search the pandas documentation using my tool.
Act 1: [PANDAS TOOL]: Read in csv as DataFrame
Observation 1: read_csv(path: str = None, …
Thought 2: Now that I found the pd.read_csv() documentation, I can answer the question.
Act 2: I know the final answer! You can find the code written below:
import pandas as pd
df = pd.read_csv('test.csv')



Sumers et al., 2023

Try Pitch

COGNITIVE ARCHITECTURES FOR LANGUAGE AGENTS



Sumers et al., 2023

Try Pitch

Section 03

DEMONSTRATION

Try Pitch

Section 04

EVALUATION

Try Pitch

EVALUATION SETUP

- **GPT-4-32k** with training data up to Sep 2021
- Test cases composed of prompt, data input, and correct function
- **20** test cases, 10 each with pandas functions before and after Sep 2021
- Zero-Shot, Zero-Shot structured, One-Shot prompting
- Automated code evaluation based on defined parameters
- Metrics: **accuracy** (output equivalency), **cost**, **time taken**

Try Pitch

TESTING FRAMEWORK

🎯 GOAL:

Compare generated solution by output code with generated solution by given code.

⚠️ CHALLENGES:

- handling string inputs
- defining logic

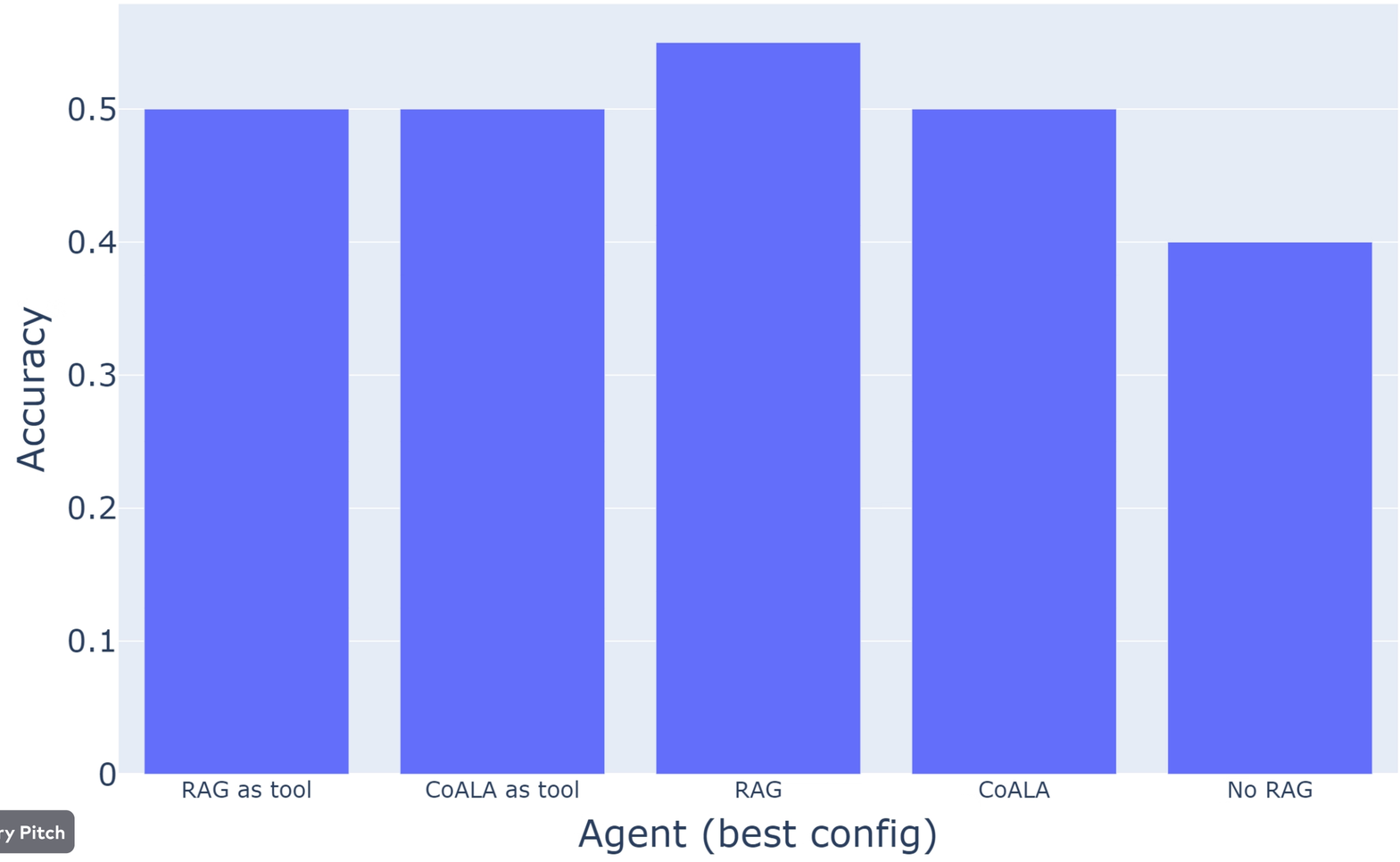
✅ SOLUTIONS:

- executing in different namespaces
- compare based on object types

Try Pitch

ZERO-SHOT PROMPTING RESULTS

Accuracy for solving test cases

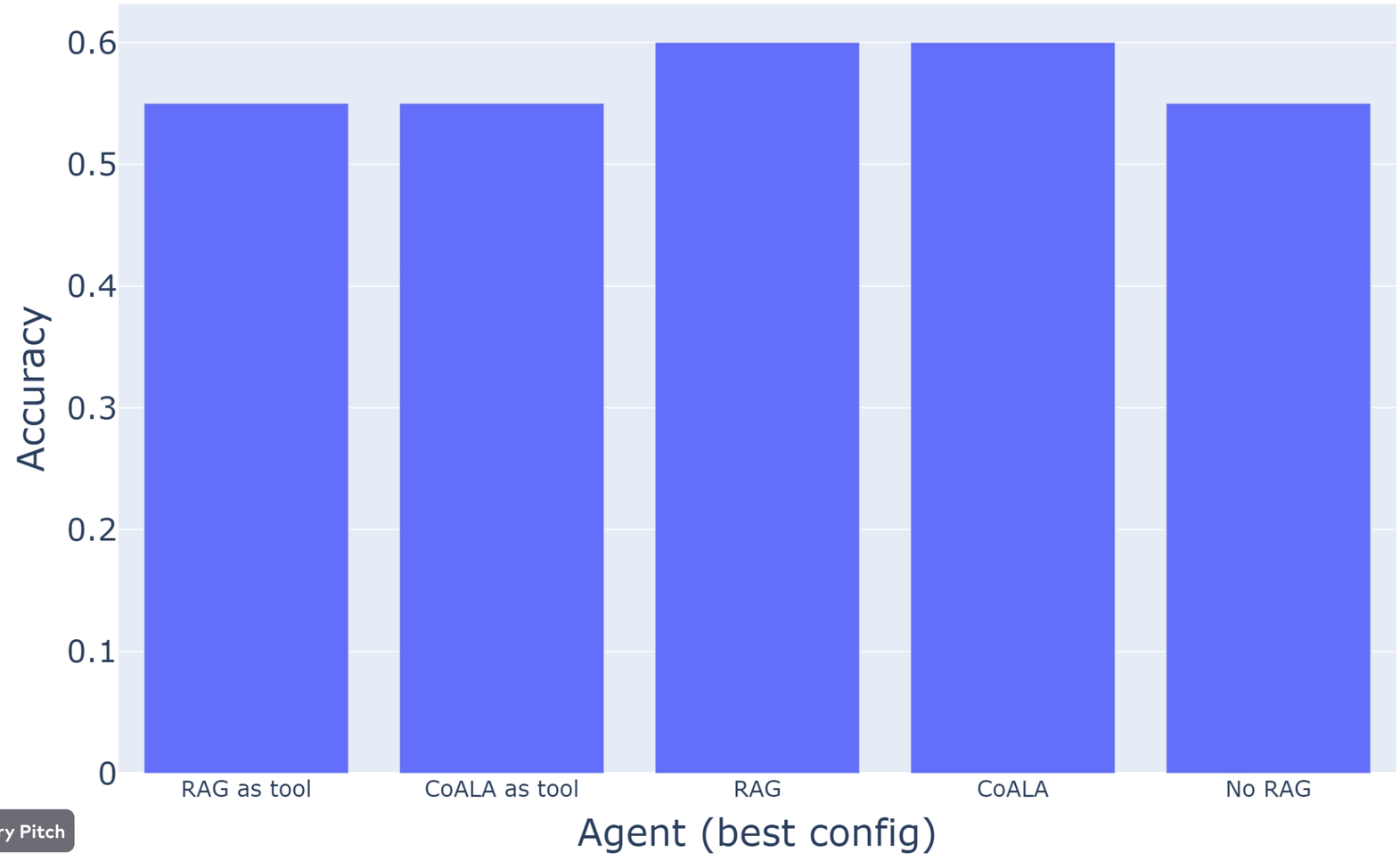


15% ACCURACY IMPROVEMENT

Try Pitch

ZERO-SHOT STRUCTURED PROMPTING RESULTS

Accuracy for solving test cases

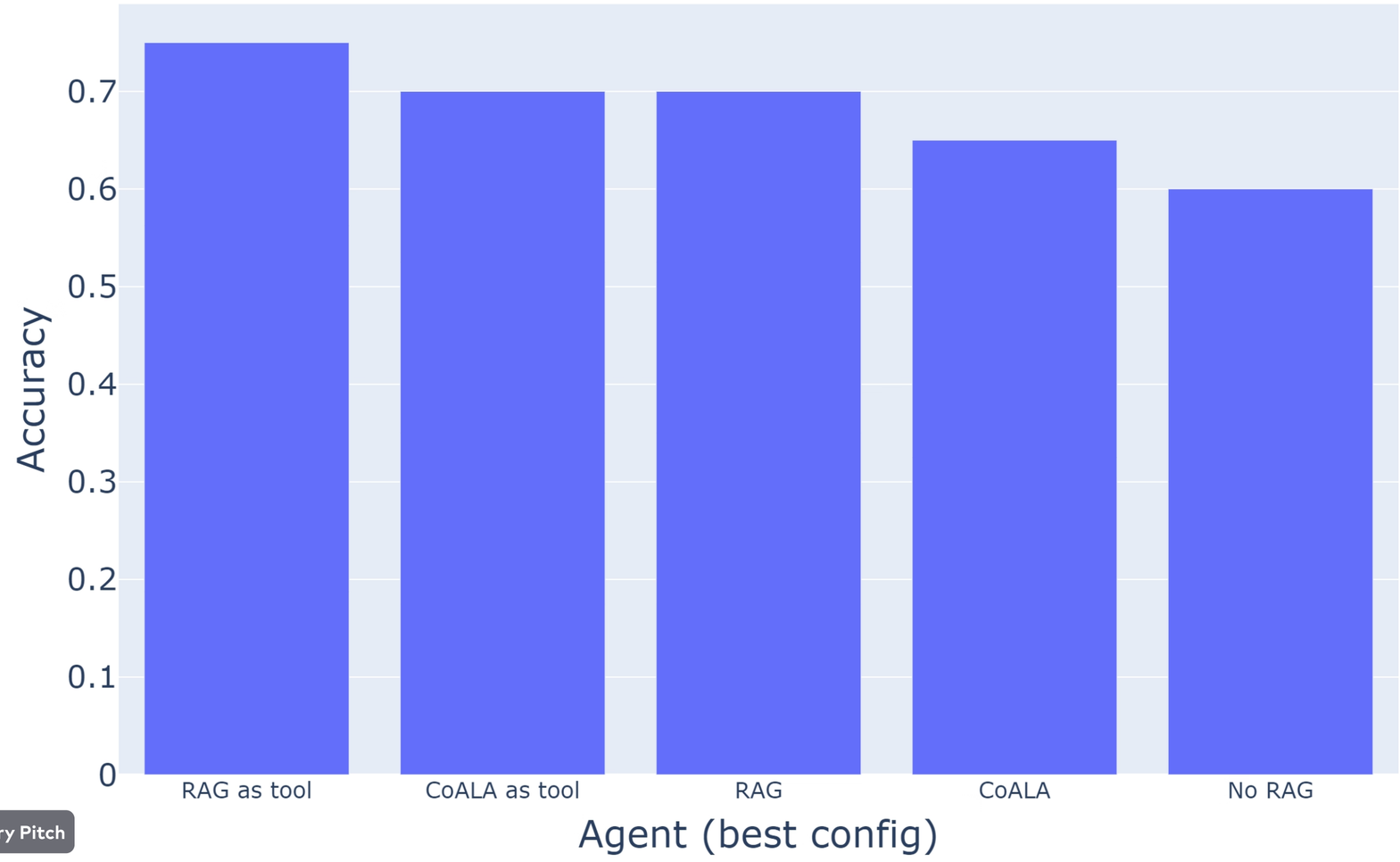


5% ACCURACY IMPROVEMENT

Try Pitch

ONE-SHOT PROMPTING RESULTS

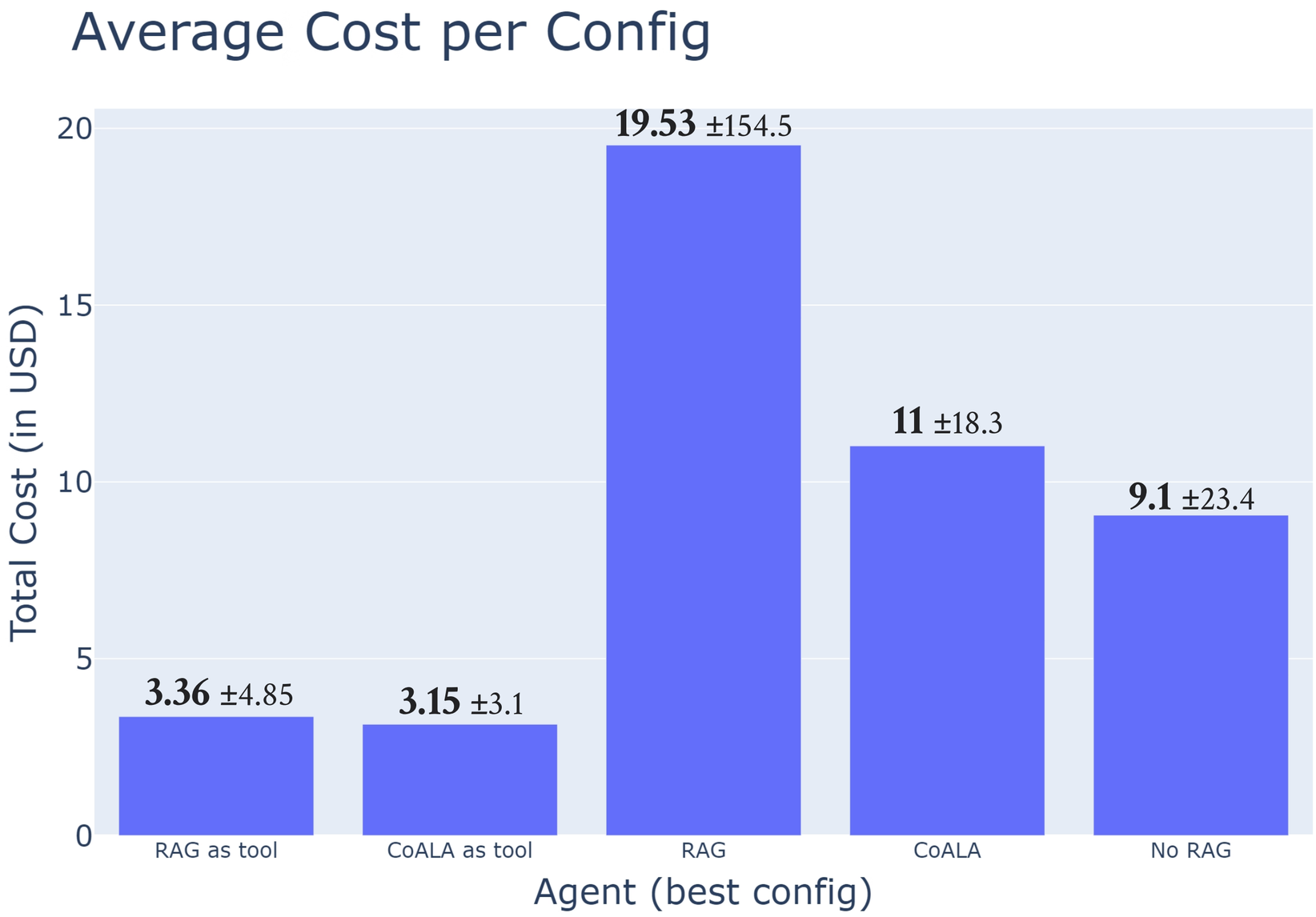
Accuracy for solving test cases



15% ACCURACY IMPROVEMENT

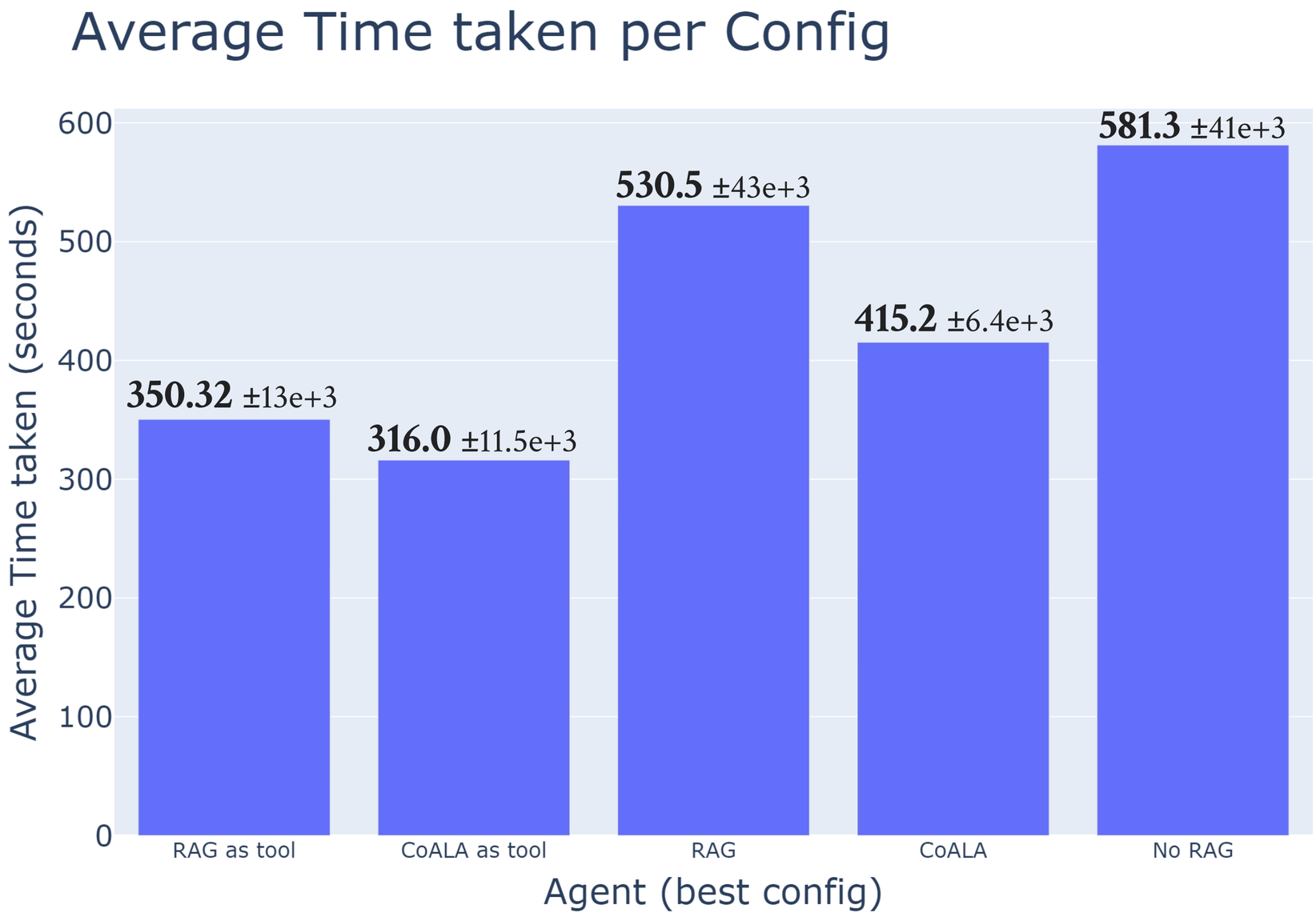
Try Pitch

EVALUATION RESULTS



Try Pitch

EVALUATION RESULTS



Try Pitch

EVALUATION TAKEAWAYS

- RAG, CoALA **improve response accuracy by ~10%**

- **Time, Cost generally (increasing)**:
  No RAG → RAG/CoALA as tool → RAG/CoALA

- Variance due to **Agent decision, unparseable code loop**

Try Pitch

OUR RECOMMENDATION

- Use Agents **incrementally**, start without RAG/CoALA
- **Evaluate** wether the significant cost/time increase is **worth the improvement in accuracy**
- **Structure prompts effectively**
  - Specifity, One question at the time, provide examples

Try Pitch

NLP Project: Enriching LLMs with RAG

THANK YOU

Jan Henrik Bertrand, André Gilbert, Felix Noll, Marc Grün

Try Pitch

CONTRIBUTIONS

AGENT, COALA, RAG

Jan Henrik Bertrand

DOCUMENTS, TEST CASES

Felix Noll

CLIENT, AUTO EVALUATION, RAG

André Anan Gilbert

TESTING FRAMEWORK, PRESENTATION

Marc Grün

Try Pitch

BIBLIOGRAPHY

Sumers, T.R., Yao, S., Narasimhan, K. and Griffiths, T.L., 2023. Cognitive Architectures for Language Agents (CoALA). arXiv preprint arXiv:2109.12394. Available at: https://arxiv.org/abs/2109.12394 [Accessed 11 Jan. 2024]

Try Pitch

Pitch

Want to make a presentation like this one?

Start with a fully customizable template, create a beautiful deck in minutes, then easily share it with anyone.

Create a presentation (It's free)