

The Socratic Method & IDA

System 1

机器学习

Supervised Learning

polynomial



判定工作



System 2

苏格拉底法

Dialogue Method

"什么都不知道"

"了解字面的什么"

做什么事情

为什么做这些事情

1. 定义

- 例子

Key Point

neighbors

Context

今年 NLP \Rightarrow Large Pre-trained Model

Attention \rightarrow 输入与输出
 \rightarrow diversify
 \rightarrow 层次 (training process 纠错机会)
depth

LLM 限制!

不是越大越好

① Mask 字 是否 在 改变
(entropy of 字)

② Lossless Compression
"Shannon Limit"

理论 bound

(每周 2-3 个 问题)

talk (什么问题 无法解决)

Context Window ≈ 8000

4000 - 8000

Memory 大

2/3

Open (Virtual Assistant) App
(Lam, Manning)

The Socratic Method

批判性思維的步驟

1. 提出開放式問題 (ChatGPT 是炒作嗎?)
2. 澄清關鍵詞彙 (定義 GAI, LLM, GPT)
3. 列舉假設的結論 (是炒作、不是炒作)
4. 提供支持性例子和證據
5. 構思從每個理由到結論的論述 (论点)
6. 挑戰每個論述、並提出反駁論點
7. 淘汰無理、不可靠的假設 (可能回到第3步，反覆進行)
8. 總結理由並得出結論
9. 反思過程

Techniques:

柏拉圖

蘇格拉底對話法--- 定義

1. 定義法 definition：澄清和解釋關鍵詞和概念的含義。
2. 歸納法 generalization：從潛在的觀察和理論模式中提取一般原則。歸納法用於形成更確定和全面的結論。(正確率高)
3. 歸納推理 induction：與歸納法類似，但歸納推理僅基於經驗證據。歸納推理提供高度不確定的假設。
4. 推論法 elenchus：涉及交叉審查 cross examination，使用一系列問題來檢驗假設 hypotheses 和信仰 beliefs 的一致性 consistency 和關連性 coherence。推論法旨在測試某人的論據的有效性，並幫助他們完善自己的思維，最終提出有充分支持的假設。
5. 假設排除法 hypothesis elimination：通過對抗事例 counterexamples 和邏輯推理來檢驗謬誤的假設和信仰。與推論法不同，假設排除法通過證據和邏輯來測試假設是否為真或假。

3/1/23

echang@cs.stanford.edu

5

蘇格拉底對話法 --- 定義

6. 助產法 maieutic：通過提問鼓勵替代觀點。人們探索自己的經驗、知識、信仰，促進自我發現、創造性寫作和創新。
7. 辯證法 dialectic：通過對話或辯論探索相反觀點，以達到對主題的更深入的理解。
8. 回憶法 recollection：此方法相信知識是先天的 innate，並且人們可以通過提問來記住他們已經知道的东西。
9. 反諷法 irony：揭露無知和虛偽，指出主張和真正理解間的差距。
10. 比喻法 analogy：通過比喻來比較和對比不同的概念，以幫助人們理解複雜的思想。

3/1/23

echang@cs.stanford.edu

6

蘇格拉底對話法 --- 範例

1. 定義法 definition：什麼是對話法？LLM 的定義？
2. 歸納法 generalization：貓與狗狗有四隻腳。（極少例外）
3. 歸納推理 induction：市場調查，最近 ChatGPT 被廣泛熱議著。
(30% 左右的人們沒聽過，譬如我的家庭醫生。不少例外)
4. 推論法 elenchus：論文答辯，確定論據的有效性和邏輯性。
5. 假設排除法：公園裡樹倒了三株！雷劈？颱風？地震？人為？

例題：最近天氣好嗎？

例題：最近缺蛋嗎？

證據 (有效性, 邏輯性)

蘇格拉底對話法 --- 範例

6. 助產法 maieutic：用提問做正向引導。譬如心理醫生詢問病人的情況，又譬如老師引導學生做腦力激盪討論議題。回答者被引導到一些尚未審查的角落或不同的視角探索議題。(guided exploration)
7. 辯證法 dialectic：用提問做反向引導。ChatGPT 真的會取代一些工作嗎？十年前許多人說放射科醫生會被 AI 取代，發生了嗎？
8. 回憶法 recollection：受限 LLM 的 attention window size。
9. 反諷法 irony：不適於與 LLM 溝通。
10. 比喻法 analogy：不適於與 LLM 溝通。

蘇格拉底對話法 --- 細微區別 subtleties

- 歸納法 generalization: 從潛在的觀察和理論模式中提取一般原則。歸納法用於形成 更確定和全面的結論。(IDA)
- 歸納推理 induction: 類似歸納法，但歸納推理僅 基於經驗證據。(IDA)
- 助產法 maieutic: 鼓勵人們探索自己的經驗、知識、信仰並探索替代觀點，進行助產法。助產法促進了自我發現、創造性寫作和創新。(IDA)
- 辯證法 dialectic: 探索相反的觀點，以達到對主題的更深入的理解。大多用於哲學、政治學、文學思辨。(IDA)
- 假設排除法 hypothesis elimination: 通過對抗事例 counterexamples 和邏輯推理來檢驗謬誤的假設和信仰。與推論法不同，假設排除法通過證據和邏輯來測試假設是否為真或假。大多用於科學研究。(IDA)

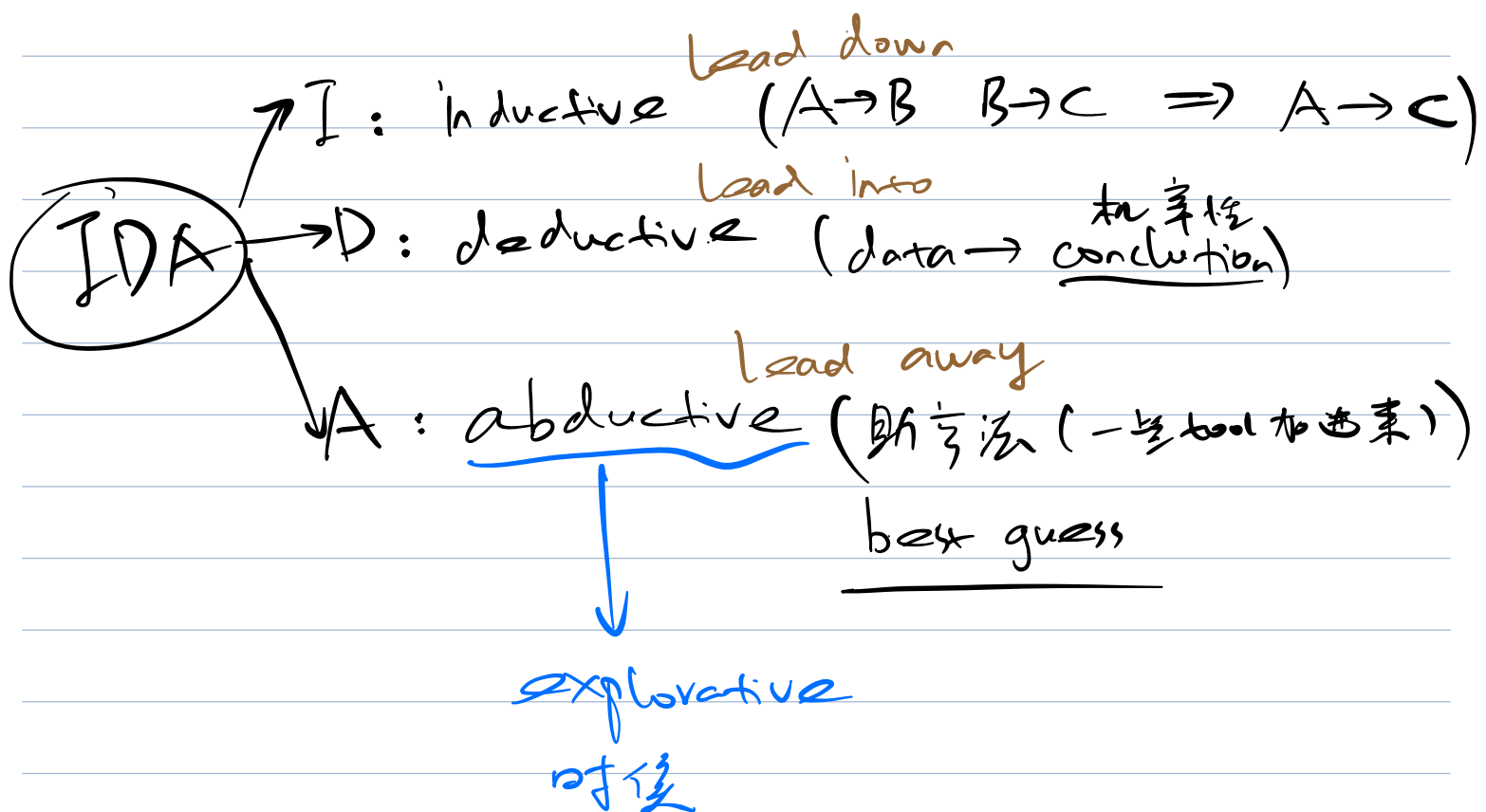
正確性
要經過
Verification

3/1/23

echang@cs.stanford.edu

9

Prompting Design: Chain of Thought



Deductive Logic

Lead *down*



Conclusion is guaranteed true

3/1/23

Inductive Logic

Lead *into*



Conclusion is probably true

echang@cs.stanford.edu

Abductive Logic

Lead *away*



Conclusion is a best guess

Kappa Online THE FORUM

Formal Reasoning IDA

• Inductive 歸納推理

- 從一堆觀察、數據，進行判斷、分類
- 準球率 < 100%

• Deductive 演繹推理

- 從一個或多個前提中推導出一個有效的結論
- 前提#1: 所有A 是 B, 前提#2: 所有 C 是 A, 結論: 因此 C 是 B
- 把握 100%

• Abductive 溯因推理

- 溯因推理做有根據的猜測或假設來解釋難以解釋的觀察結果或現象。
- 溯因推理立定假設 hypotheses 與解釋，必須後續認證或否決。
- 通常是誤判，把握低。

3/1/23

echang@cs.stanford.edu

13

Inductive Reasoning 歸納推理

- 從一堆觀察、數據，進行判斷、分類
- 準球率 $< 100\%$
- 監督學習 supervised learning 是最好的例子
- 例子，狗狗有四隻腿, $p(\text{狗} \mid \text{動物有四隻腿}) < 1$
- 一般方法：
 - Machine learning,
 - Statistical analysis (regression, correlation),
 - Bayesian models (probabilistic), and
 - Causal strength

Not scale

beach \rightarrow traffic jam

Implementation
Not scale

3/1/23

echang@cs.stanford.edu

14

Deductive Reasoning 演繹推理

- 從一個或多個前提中推導出一個有效的結論
- 演繹邏輯三段論
 - 前提1：所有 A 都是 B
 - 前提2：C 是 A
 - 結論：因此，C 是 B
- 例如，考慮以下三段論：
 - 前提1：所有人都會死亡
 - 前提2：蘇格拉底是人類
 - 結論：因此，蘇格拉底會死亡

3/1/23

echang@cs.stanford.edu

15

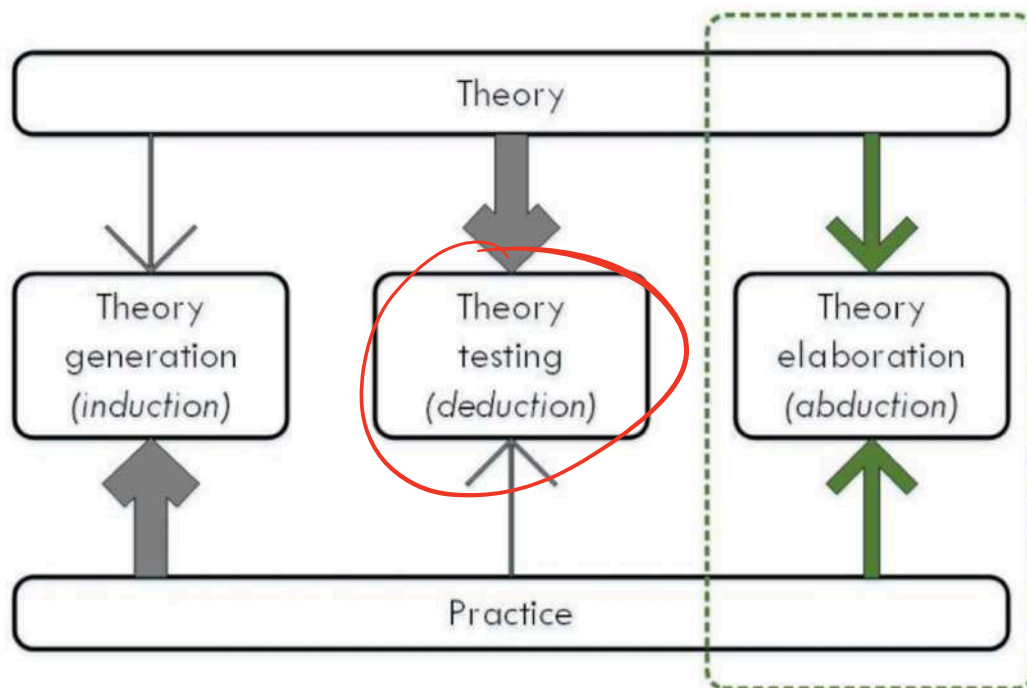
Abductive Reasoning 從屬推理

- 使用可用的證據和知識為給定的觀察或一組觀察形成可能的最佳解釋，即使解釋不一定確定或完整。
- 例 #1: 醫學診斷。當患者出現一些 symptoms，醫生使用溯因推理對這些病症的根本原因做出有根據的臆測。醫生會考慮譬如患者病例、家族病史、最近的流行病，生成關於觀察到的現象的原因的假設，或解釋異常結果。
- 例 #2: AI 研究觀察 ChatGPT 生成意想不到的結果，根據他們的知識和現有證據，使用從屬推理來臆測關於意想不到結果的根本原因的假設。（谷歌有工程師臆測 LLM 有意識，被解雇了）
- 根據現有證據和知識做出有根據的猜測或假設，然後檢驗該假設是否可以證實或推翻它。溯因推理是許多領域的重要工具，包括科學、醫學和工程。

3/1/23

echang@cs.stanford.edu

16



3/1/23

echang@cs.stanford.edu

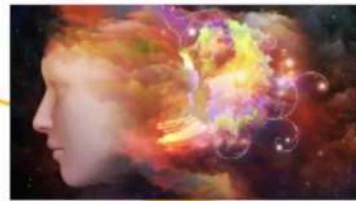
17

蘇格拉底 相會 IDA

1. 定義法 definition
2. 歸納法 generalization
3. 歸納推理 induction
4. 推論法 elenchus
5. 假設排除法
6. 助產法 maieutic
7. 辯證法 dialectic

Inductive Reasoning

Deductive Reasoning



Abductive Reasoning

3/1/23

echang@cs.stanford.edu

18

Prompting theoretical foundation

Critical Reading:

批判性閱讀
 $\Gamma = \text{CRIT}(d)$

Γ 評分 d 文章

Function $\Gamma = \text{CRIT}(d)$

Input. d : document; **Output.** Γ : validation score;
Vars. Ω : claim; R & R' : reason & counter reason set;
Subroutines. $\text{Claim}()$, $\text{FindDoc}()$, $\text{Validate}()$;

Begin

- #1 Identify in d the claim statement Ω ;
- #2 Find a set of supporting reasons R to Ω ;
- #3 For $r \in R$ eval $r \Rightarrow \Omega$
 If $\text{Claim}(r)$, $(\gamma_r, \theta_r) = \text{CRIT}(\text{FindDoc}(r))$;
 else, $(\gamma_r, \theta_r) = V(r \Rightarrow \Omega)$;
- #4 Find a set of rival reasons R' to Ω ;
- #5 For $r' \in R'$, $(\gamma_{r'}, \theta_{r'}) = V(r' \Rightarrow \Omega)$ eval rivals;
- #6 Compute weighted sum Γ , with $\gamma_r, \theta_r, \gamma_{r'}, \theta_{r'}$.
- #7 Analyze the arguments to arrive at the Γ score.
- #8 Reflect on and synthesize CRIT in other contexts.

End

PostModem

- CRIT is now for critical reading on essays, e.g., news and research.
- For literature, such as novel and poetry, one should derive a different template.
- Submitted one prompt after another produces detailed explanations and better for education purpose.
- Submitted batch or one-by-one score similarly in an SAT like test, 97.5% accuracy.

Intelligent Device Agent

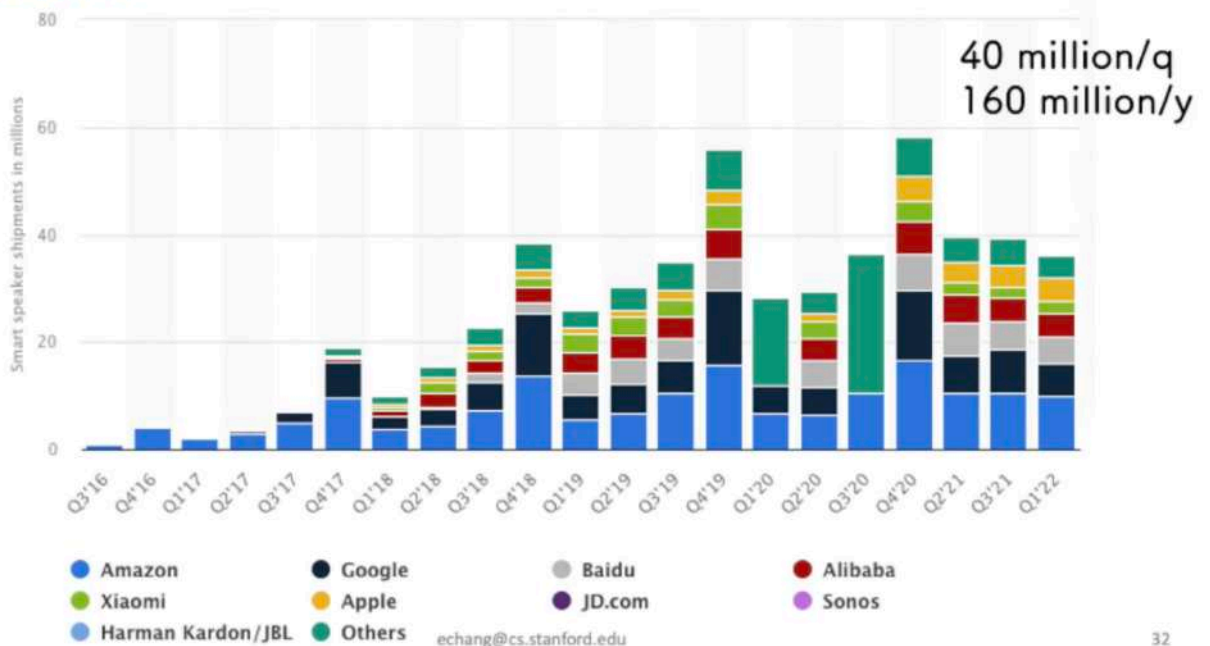


Timeline of Mainstream Voice Assistants



Smart Speaker Unit Shipments Worldwide

©Statista 2023

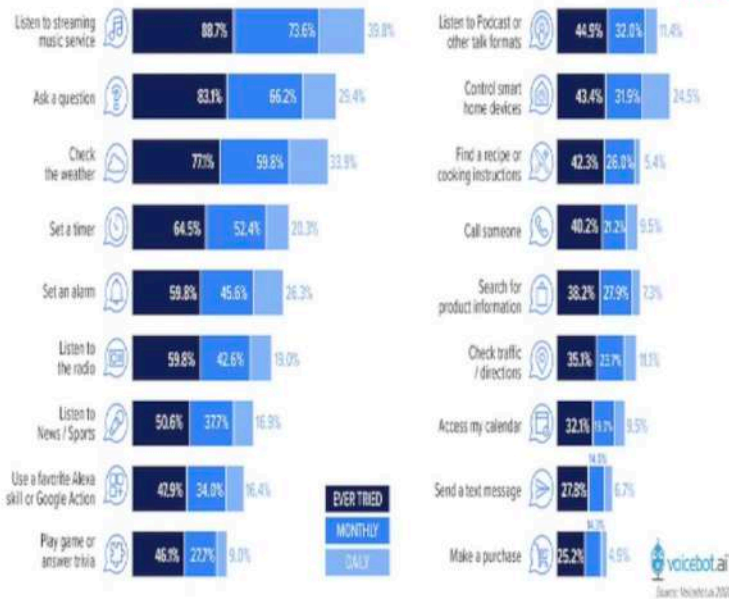


News #7 skill

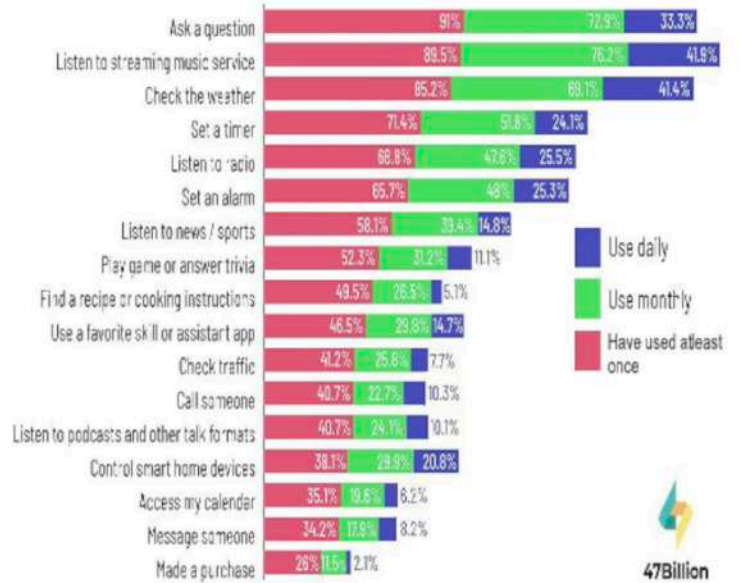
37% monthly use

15% daily use

Smart Speaker Use Case Frequency January 2020



Smart Speaker Use Case Frequency



Andrew Ng's tweet, December 2016

2016

"As speech-recognition accuracy goes from 95% to 99%, we'll go from barely using it to using all the time!"

2016 is 3 years before the birth of BERT!

2016 → 2019 GPT2

2020 → GPT3 & GSHard

2022 → ChatGPT

→ Not only \$, knowledge
Virtual Assistant

你要 LLM 幫你做什麼？ 你如何幫 LLM 準備？



UI: ChatGPT

Large Language Model(s): e.g., GPT3

Attention Mechanism, polysemy w/ context

Word2vec, similarity

Knowledge, represented by a huge set of documents

3/1/23

echang@cs.stanford.edu

35

你要 LLM 幫你做什麼？ 你如何幫 LLM 準備？



UI: ChatGPT

Large Language Model(s): e.g., GPT3

Attention Mechanism, polysemy w/ context

Word2vec, similarity

Knowledge, represented by a huge set of documents

3/1/23

echang@cs.stanford.edu

36

你要 LLM 幫你做什麼？ 你如何幫 LLM 準備？



Dialogues

Natural Language

UI: ChatGPT

Actions: input & output

Natural Language

Large Language Model(s): e.g., GPT3

3/1/23

echang@cs.stanford.edu

37

ChatGPT → 虛擬助理

- 了解使用者目的 intent (要什麼? 旅行)
- 獲取情境 context (目的地? 時間? 機票? 旅館? 租車? Osaka, 五月, 一周)
- 了解 LLM 的能力 skills
 - 聊天
 - 自然語言處理
 - 會訂機票, 訂房, 租車嗎?
- LLM 無這些能力, 必須提升到虛擬助理層級
 - 訂機票 APIs, e.g., Reserve-Flights(from, to, departure, return date, class, etc.)
 - 訂房 APIs, e.g., Reserve-Lodging(city, brand, price, check-In/out, room type, etc.)
 - 租車 APIs

3/1/23

echang@cs.stanford.edu

39

LLM \Rightarrow VA

LLM → Virtual Assistant (VA) 虛擬助理

Large Language Model



3/1/23

echang@cs.stanford.edu

40

你要 VA 幫你做什麼？ 你如何幫 VA 準備？



Dialogues

Natural Language

VA

Natural Language

Actions: input & output

Programming Language

LLMs

Schema.org

APPs

IoT

.....

3/1/23

echang@cs.stanford.edu

41

Stanford OVAL Thingpedia: Encyclopedia of Things

stores knowledge about devices in the world



- IoT Device
- Attributes
- API (Function) and its Corresponding NL

	Natural Language	API Signatures
WHEN	@Stanford tweets	Monitor (@home_timeline(), ...) author=="Stanford")
GET	tweets matching "#Cardinal"	search(...), contains (hashtag, ...)
DO	tweet "Stanford won!"	post (status)

STANFORD



Brad's Travel Arrangement

- Make relevant reservations
 - Monitor schedule and automatic alert and rebook
- Manage electric devices at home when traveling
- Monitor home safety, e.g., cameras and smoke detectors
- Check apps
- Schedule meetings
- Etc.



API

Stanford Open Virtual Assistant Lab

Senior Members



Prof. Monica Lam (PI, CS)



Prof. Michael Bernstein (CS)



Prof. Dan Boneh (CS)



Prof. Edward Y. Chang (CS, Adjunct)



Dr. Jen King (Law)



Prof. James Landay (CS)



Prof. Fei-Fei Li (CS)



Prof. Chris Manning (CS)



Prof. David Mazieres (CS)



Prof. Chris Re (CS)



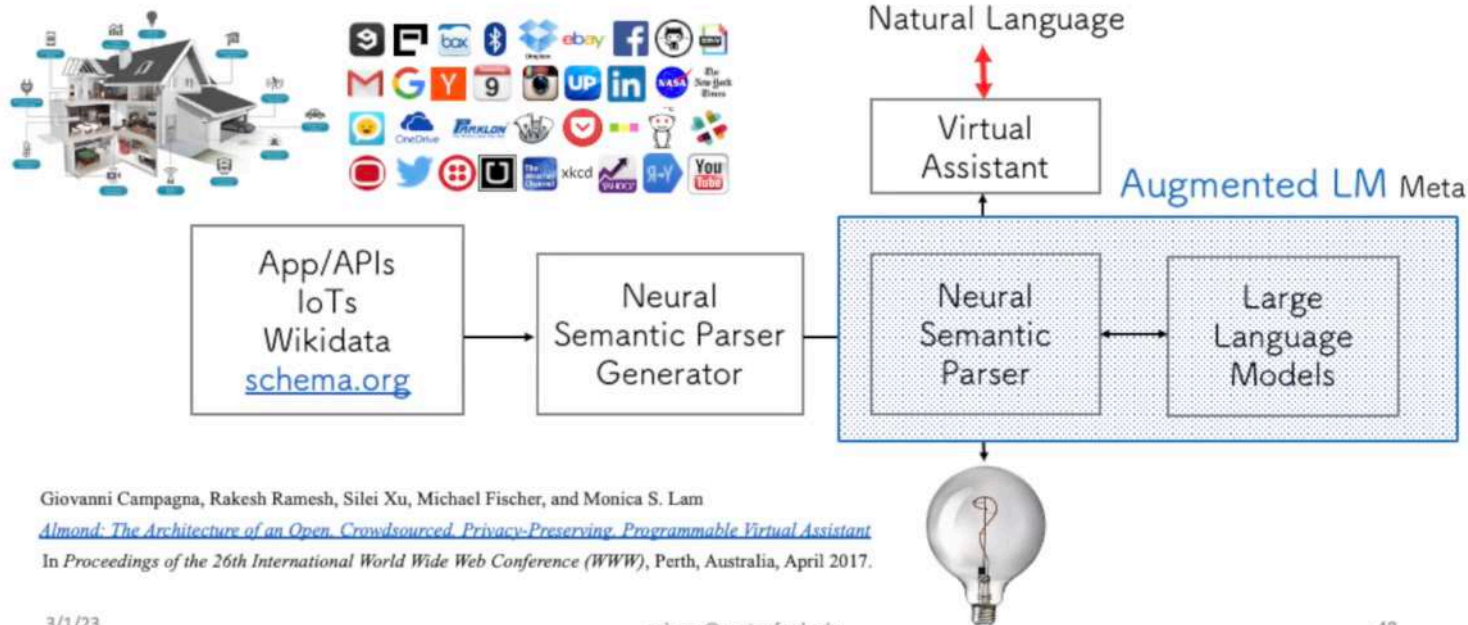
Dr. Richard Socher

3/1/23

echang@cs.stanford.edu

46

OVAL Components



Giovanni Campagna, Rakesh Ramesh, Silei Xu, Michael Fischer, and Monica S. Lam

Almond: The Architecture of an Open, Crowdsourced, Privacy-Preserving, Programmable Virtual Assistant

In Proceedings of the 26th International World Wide Web Conference (WWW), Perth, Australia, April 2017.

3/1/23

echang@cs.stanford.edu

48

6d



Natural Language

Virtual Assistant

Neural Semantic Parser

Large Language Models

Programming Language



Natural Language

Virtual Assistant

Neural Semantic Parser

Large Language Models

Programming Language



3/1/23

49

Example Dialogue Grounding

- User: “冷的時候開暖氣”
- VA performs NLU and then grounds semantic to code
 - ⊗ 開? 同義詞：打開
 - ⊗ 冷? 溫度計低於幾度? 華氏攝氏?
- 自然語言翻成程式語言 (代碼)

If (@thermometer.temperature(type=c) < 10°C) → @heater.state(on)



Key References

- Prompting Large Language Models With the Socratic Method, Edward Y. Chang, IEEE CWCC, March 2023. (first version January 25th) [\[link\]](#)
- CRIT: An Inquisitive Prompt Template for Critical Reading, Edward Y. Chang, Computer Science, Stanford University, February 2023. [\[link\]](#)
- About IDA [\[link\]](#)

