Converge to the Truth:
Factual Error Correction via Iterative Constrained Editing

Jiangjie Chen[1,*] (jjchen19@fudan.edu.cn), **Rui Xu**[1,*], **Wenxuan Zeng**[2], **Changzhi Sun**[3], **Lei Li**[4], **Yanghua Xiao**[1]

[1]Fudan University [2]University of Electronic Science and Technology of China [3]ByteDance AI Lab [4]University of California, Santa Barbara
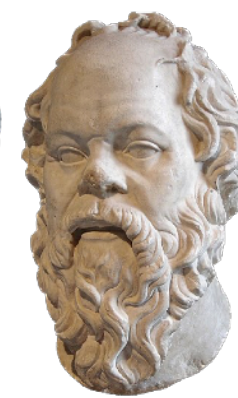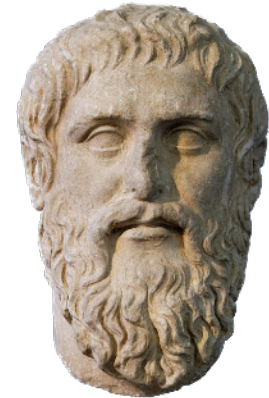
Project Page

Factual Error Correction

How to correct a sentence with factual errors with trustworthy knowledge bases?

- ✘ James Cameron directed **Thor 2**, which was released in 2022.
- ☑ James Cameron directed Avatar 2, which was released in 2022.

- ✘ **Socrates** wrote the Ethics and the Republic.
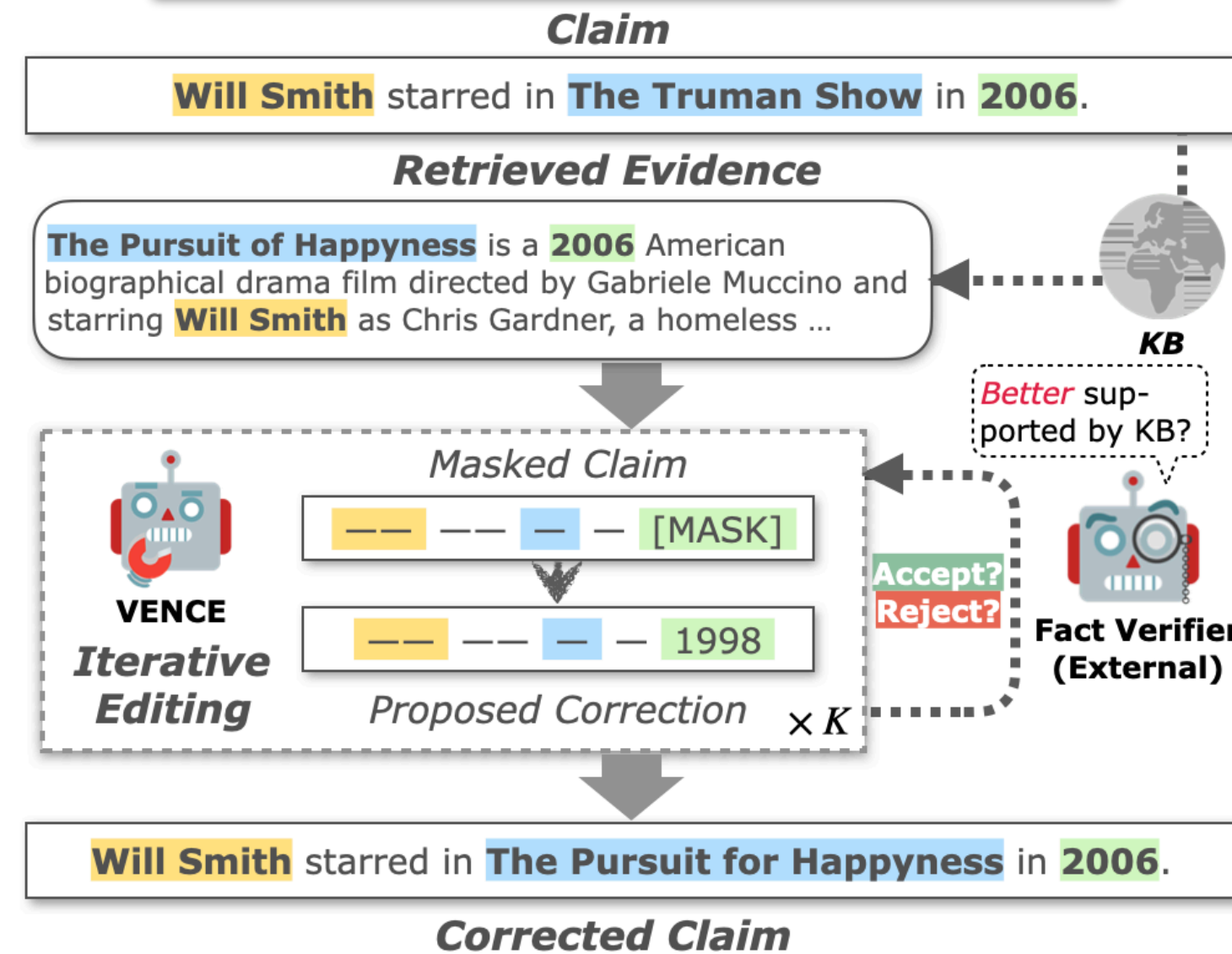- ☑ Platos wrote the Ethics and the Republic.

Motivations

- 📝 *Over-erasure:*
  - ✦ Correct errors via iterative editing
- 🕵️ *Invalidated Revision:*
  - ✦ Fact Verification + Factual Error Correction
- 🧐 *Lacking Supervision:*
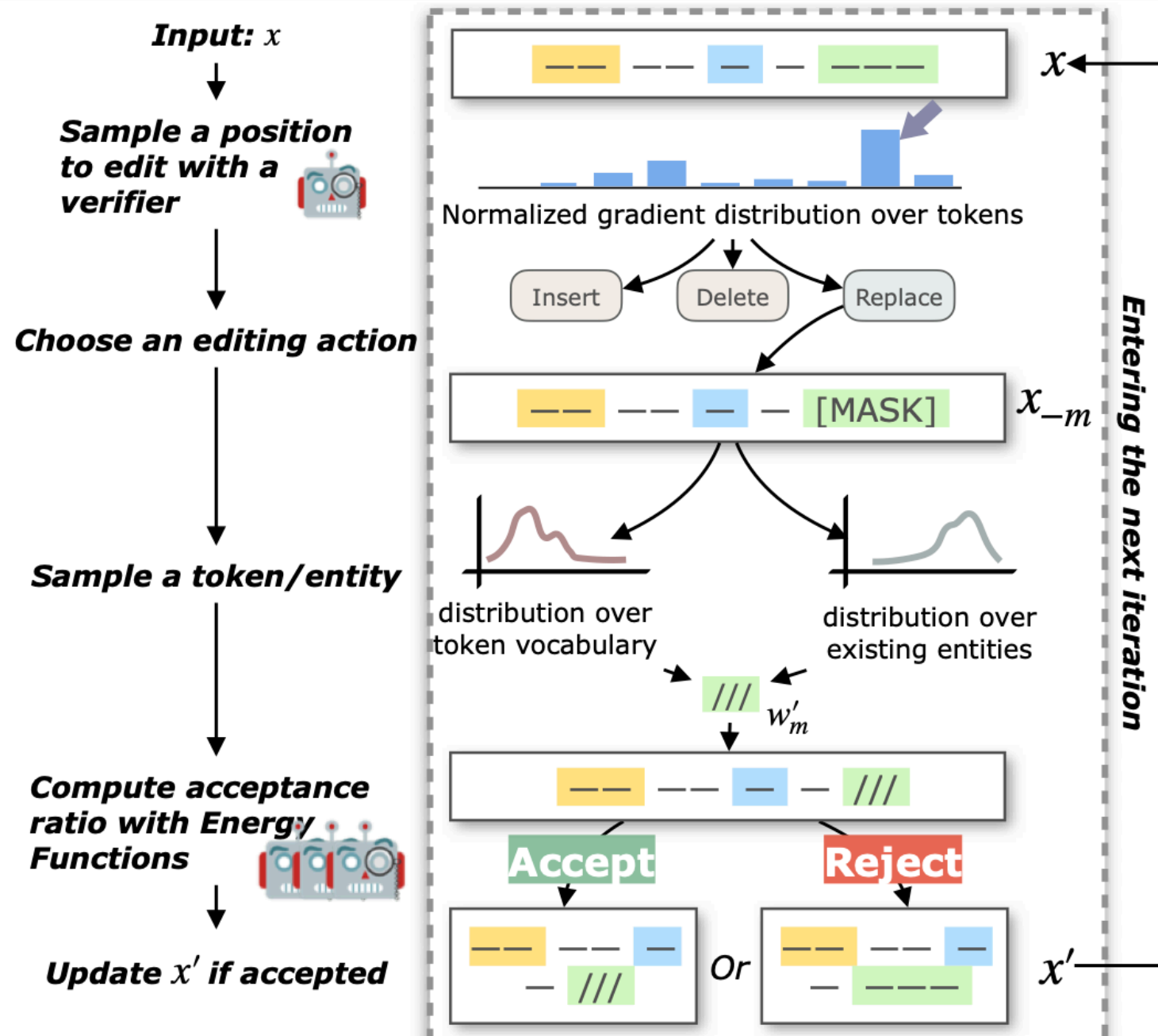 - ✦ Distant supervision from Fact Verification Helps!

Contributions

- ❖ *Iteratively* correcting factual errors;
- ❖ *No need* for direct supervision;
- ❖ Make *minimal* but *necessary* revisions;
- ❖ Effective integration of external *fact verification* signal.

The VENCE Framework

Claim

Will Smith starred in **The Truman Show** in **2006**.

Retrieved Evidence

The Pursuit of Happyness is a **2006** American biographical drama film directed by Gabriele Muccino and starring **Will Smith** as Chris Gardner, a homeless ...

KB

Better supported by KB?

VENCE Iterative Editing

Masked Claim

— — — — — [MASK]

Proposed Correction   × K

— — — — — 1998

Accept? Reject? **Fact Verifier (External)**

Will Smith starred in **The Pursuit for Happyness** in **2006**.

Corrected Claim

Workflow

Input: x

Sample a position to edit with a verifier

Normalized gradient distribution over tokens

[Insert] [Delete] [Replace]

Choose an editing action

— — — — — [MASK] x_{-m}

Sample a token/entity

distribution over token vocabulary   distribution over existing entities

/// w'_m

Compute acceptance ratio with Energy Functions

Accept   **Reject**

— — — — ///   Or   — — — — — x'

Entering the next iteration

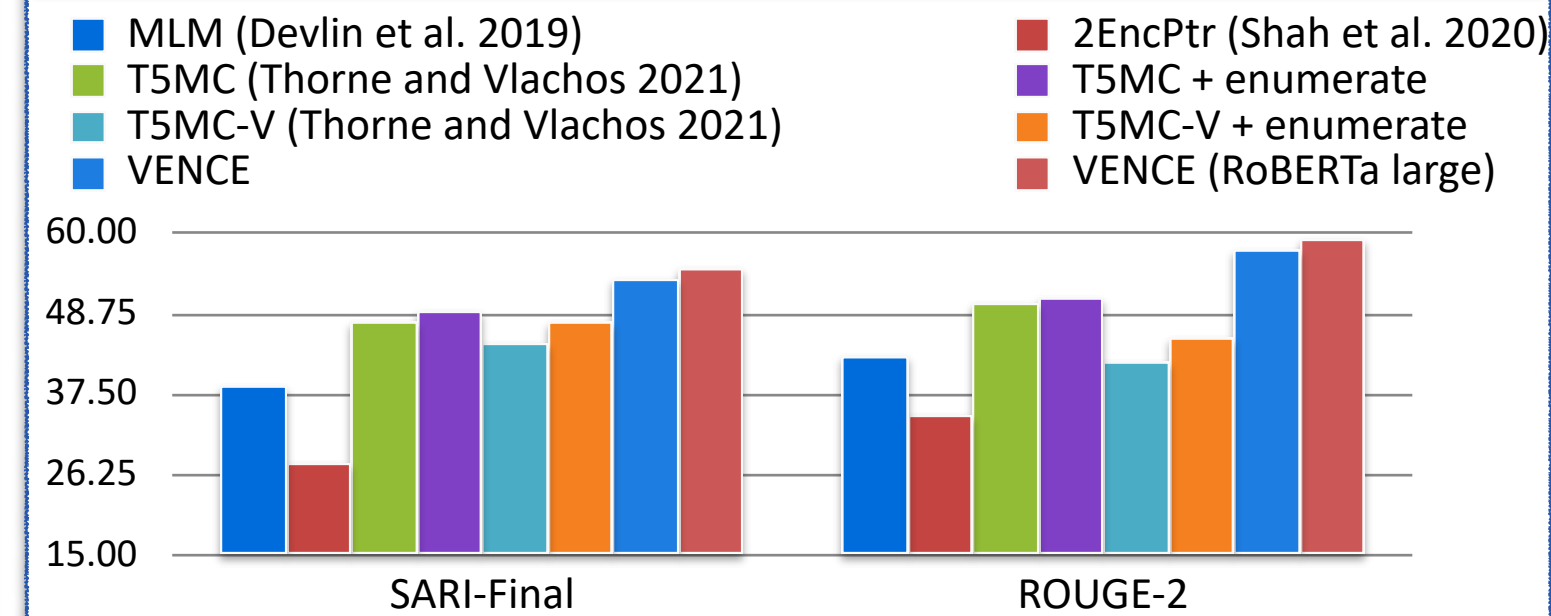**Update** x' **if accepted**

Experiments

Dataset
- ❖ **FECData** (Thorne and Vlachos 2021)
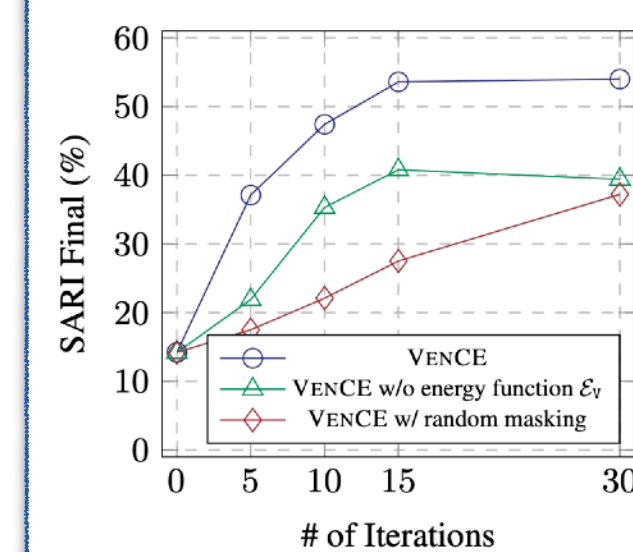- ❖ **FEVER** (Thorne et al. 2018)

Metrics
- ❖ **SARI scores**: F1 of words being added/deleted/kept
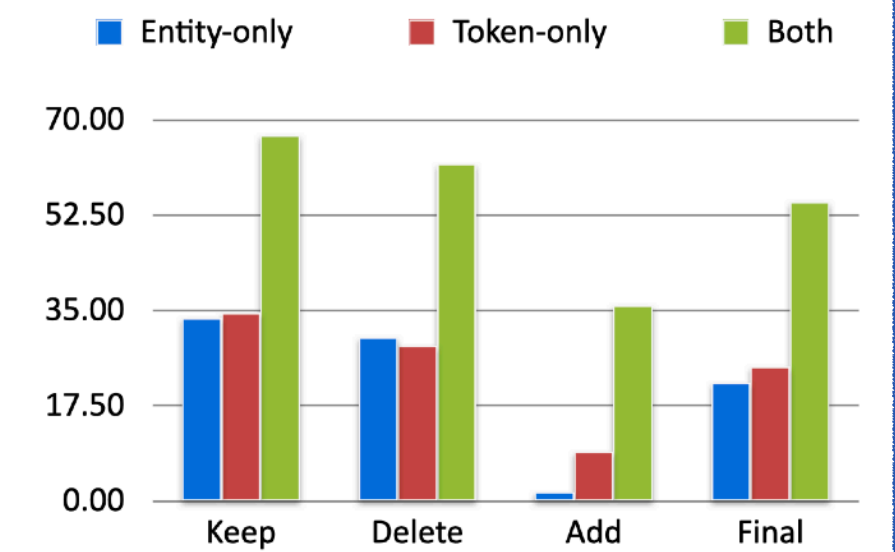- ❖ **Human scores**

Introducing VENCE: The state-of-the art distant supervised FEC method.

Legend:
- MLM (Devlin et al. 2019)
- T5MC (Thorne and Vlachos 2021)
- T5MC-V (Thorne and Vlachos 2021)
- VENCE
- 2EncPtr (Shah et al. 2020)
- T5MC + enumerate
- T5MC-V + enumerate
- VENCE (RoBERTa large)

(Bar chart: SARI-Final, ROUGE-2; y-axis 15.00 to 60.00)

VENCE converges, and faster.

(Line chart: SARI Final (%) vs # of Iterations; legend: VENCE, VENCE w/o energy function $\mathcal{E}_T$, VENCE w/ random masking)

Proposal from both entity & token spaces matters!

Legend: Entity-only, Token-only, Both

(Bar chart: Keep, Delete, Add, Final; y-axis 0.00 to 70.00)

According to Humans: VENCE productions are not only supported, but with errors corrected.

Legend: T5MC, T5MC-V, VENCE

(Bar chart: Grammatical, Supported, Corrected; y-axis 45.00 to 100.00)