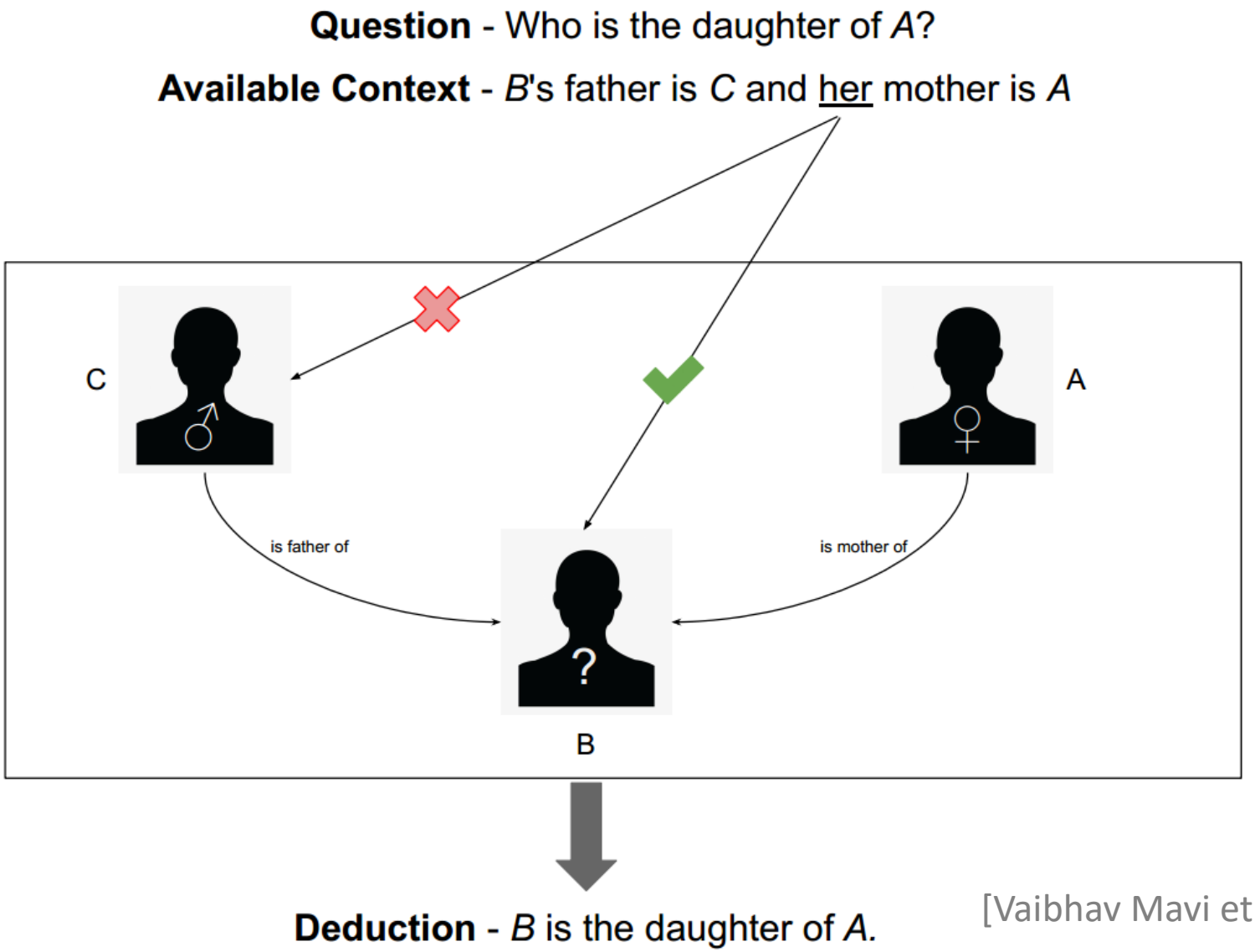
Prompting Explicit and Implicit Knowledge for Multi-hop Question Answering Based on Human Reading Process

Guangming Huang, Yunfei Long, Cunjin Luo, Jiaxing Shen, Xia Sun

Multi-hop QA



[Vaibhav Mavi et la, 2022]

Summary of our work

- **An effective approach based on human reading process**
- **Comparable performance with SOTA on HotpotQA**
- **Ablation studies supports our hypothesis for our proposed model, grounded in the human reading process.**

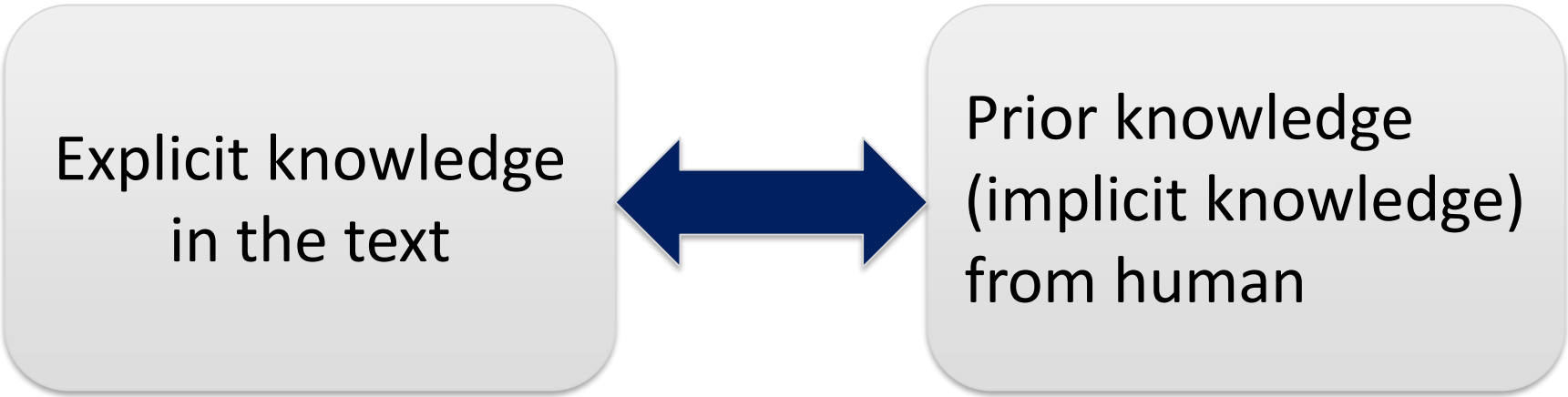
Motivation and Theories:

Human reading comprehension studies

- **Information sources are often repeated, leading to redundancy** [Smith, 1971]
- **Readers can reduce their reliance on explicit information** [Hagoort et al., 2004]
- **Readers both use explicit information and their prior knowledge** [Clarke and Silberstein, 1977]
- **Experimental results show a connection between the prior knowledge and reading comprehension** [Abdelaaland Sase, 2014]

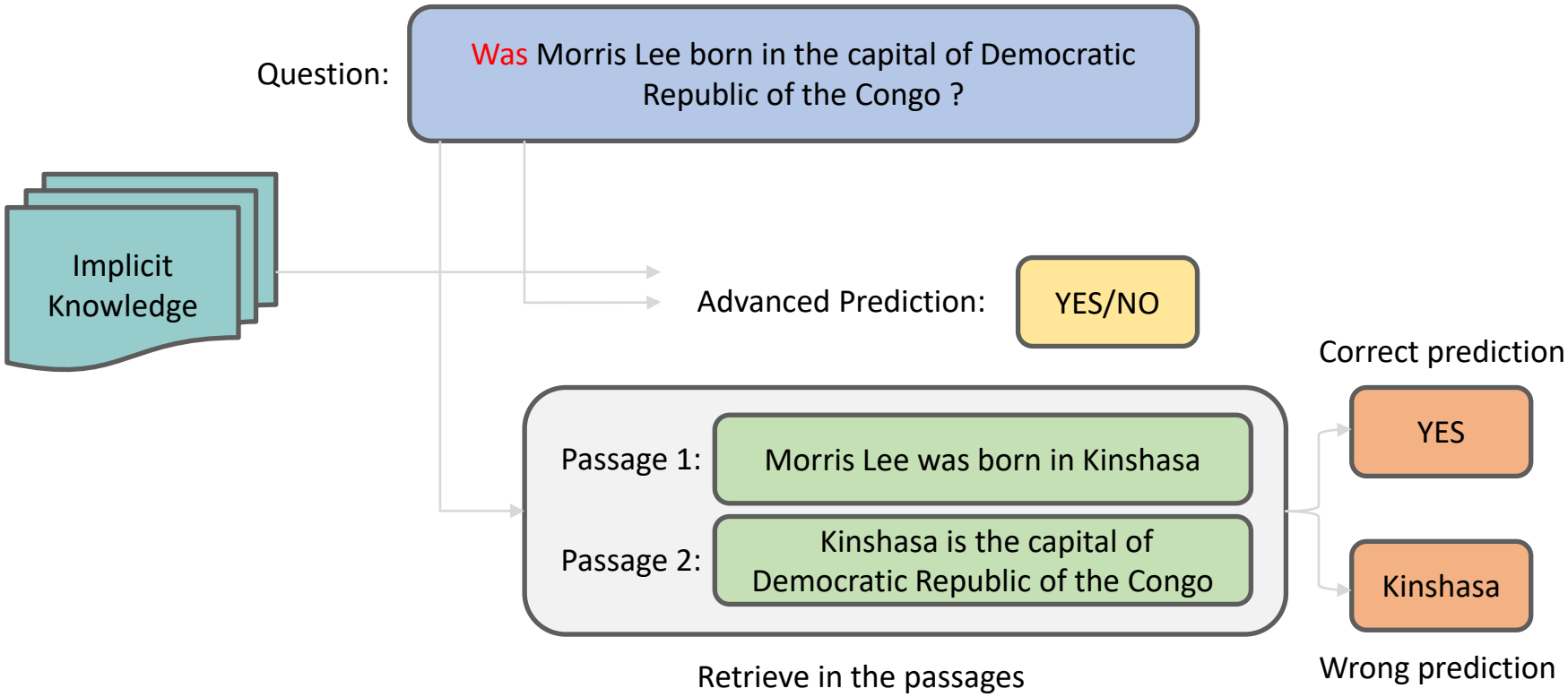
Motivation and Theories:

Human reading comprehension studies



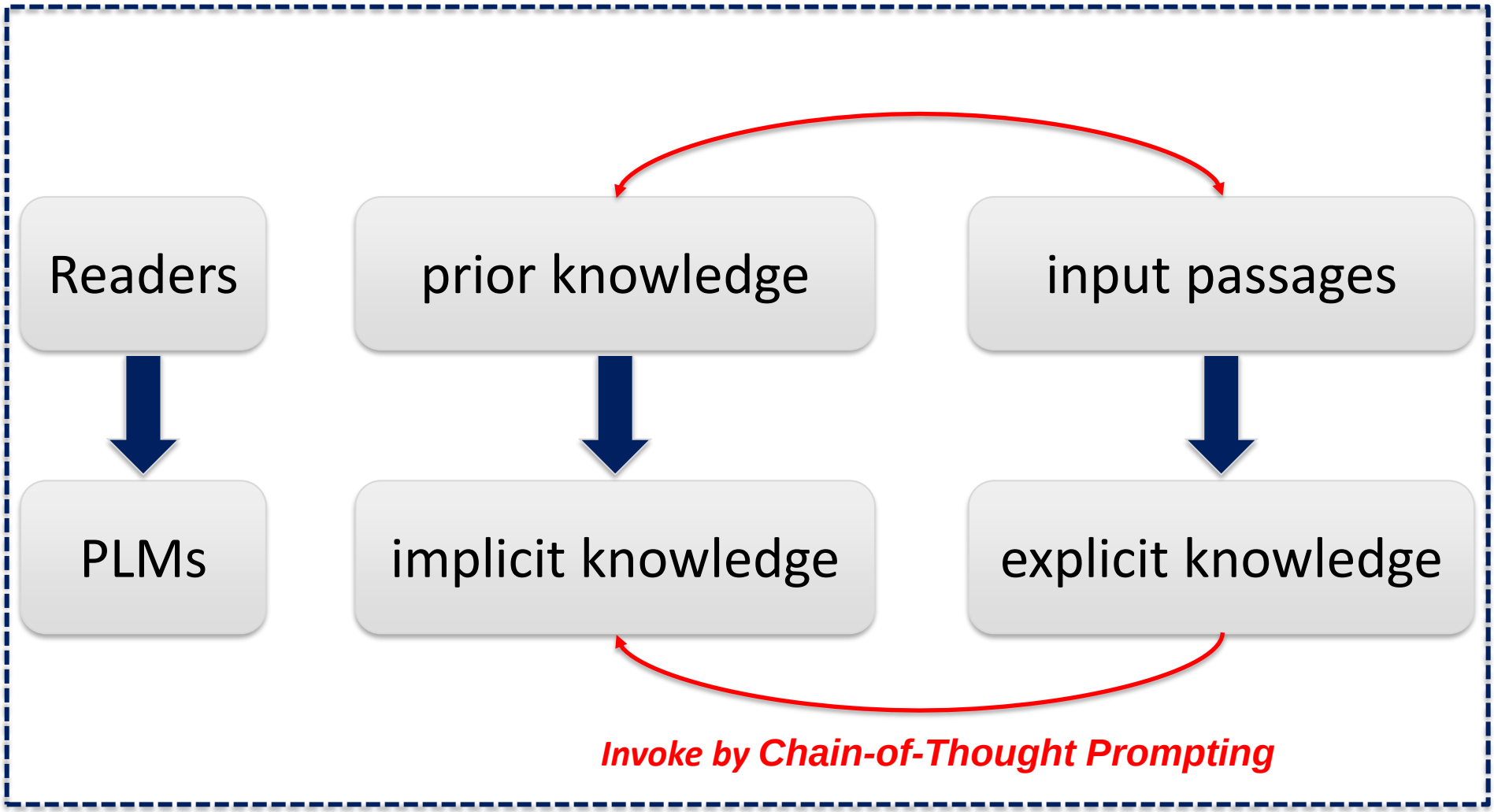
Motivation and Theories:

An example of the significance of implicit knowledge in reading comprehension.

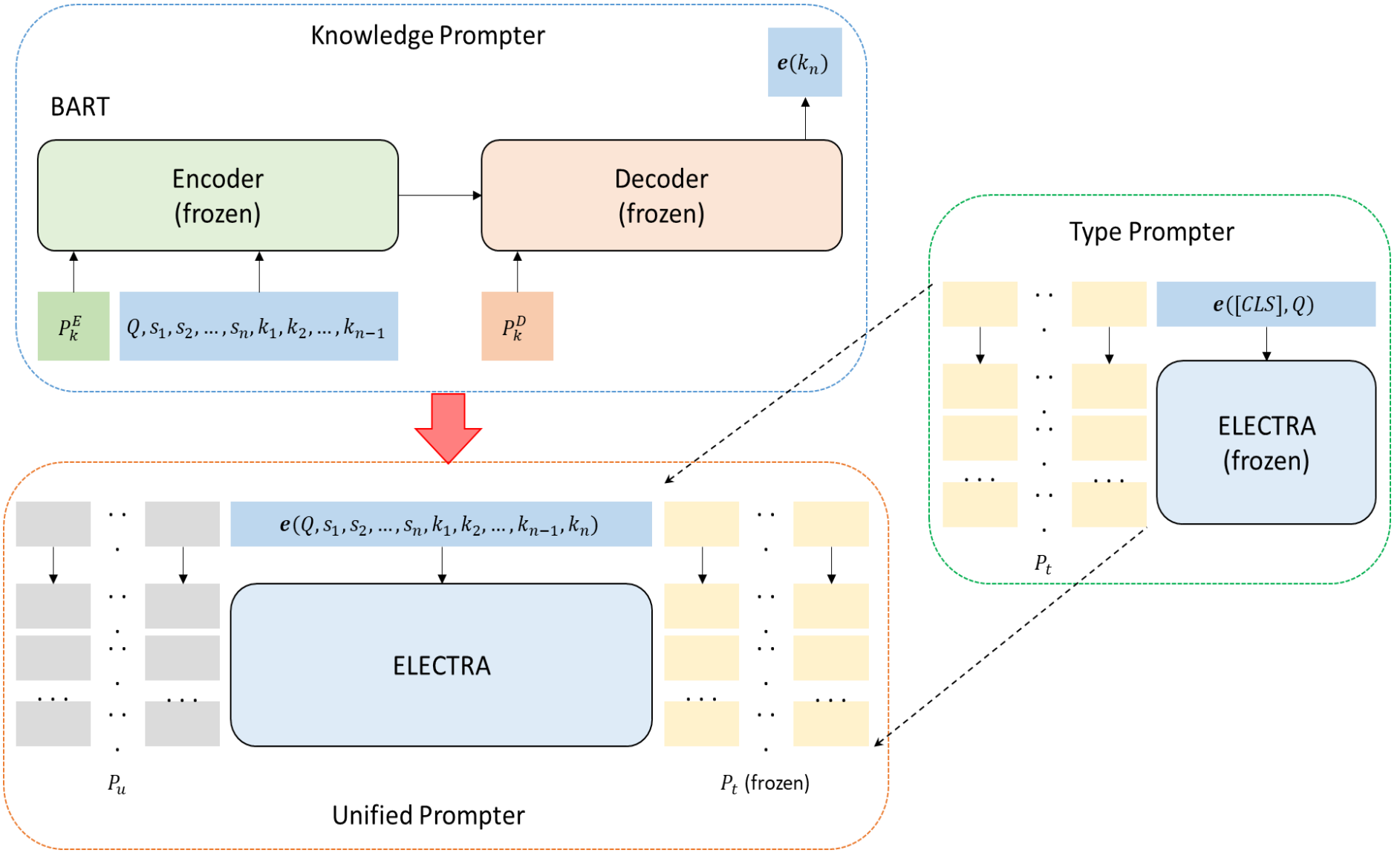


Methodology

Aligning with human reading process



Methodology: PEI framework



Methodology: PEI framework

An iterative approach

$$P\left(k_j \middle| Q, S_j, K_{j-1}\right) = \prod_{j=1}^{n} P\left(k_j \mid Q, s_1, \ldots, s_j, k_1, \ldots, k_{j-1}\right) \quad (1)$$

$$decoder\left(k_j\right) = encoder\left(Q, S_j, K_{j-1}\right) \quad (2)$$

$$decoder\left(k_1\right) = encoder\left(Q, s_1\right) \quad (3)$$

Results: main results on HotpotQA

On the blind test set of HotpotQA in the distractor setting

Models	Ans		Sup		Joint	
	EM	F1	EM	F1	EM	F1
Baseline Model (Yang et al., 2018)	45.60	59.02	20.32	64.49	10.83	40.16
DecompRC (Min et al., 2019)	55.20	69.63	-	-	-	-
OUNS (Perez et al., 2020)	66.33	79.34	-	-	-	-
QFE (Nishida et al., 2019)	53.86	68.06	57.75	84.49	34.63	59.61
DFGN (Qiu et al., 2019)	56.31	69.69	51.50	81.62	33.62	59.82
SAE-large (Tu et al., 2020)	66.92	66.92	61.53	86.86	45.36	71.45
C2F Reader (Shao et al., 2020)	67.98	81.24	60.81	87.63	44.67	72.73
Longformer (Beltagy et al., 2020)	68.00	81.25	63.09	88.34	45.91	73.16
HGN (Fang et al., 2020)	69.22	82.19	62.76	88.47	47.11	74.21
AMGN (Li et al., 2021)	70.53	83.37	63.57	88.83	47.77	75.24
S2G (Wu et al., 2021)	70.72	83.53	64.30	88.72	48.60	75.45
iCAP † (Wang et al., 2022)	68.61	81.82	62.80	88.51	47.02	74.11
PCL (Deng et al., 2022)	71.76	84.39	64.61	89.20	49.27	76.56
Beam Retrieval (Zhang et al., 2023)	72.69	85.04	**66.25**	**90.09**	**50.53**	77.54
PEI (Ours)	**72.89**	**85.32**	65.03	89.81	49.91	**77.84**

PEI achieves comparable performance with SOTA

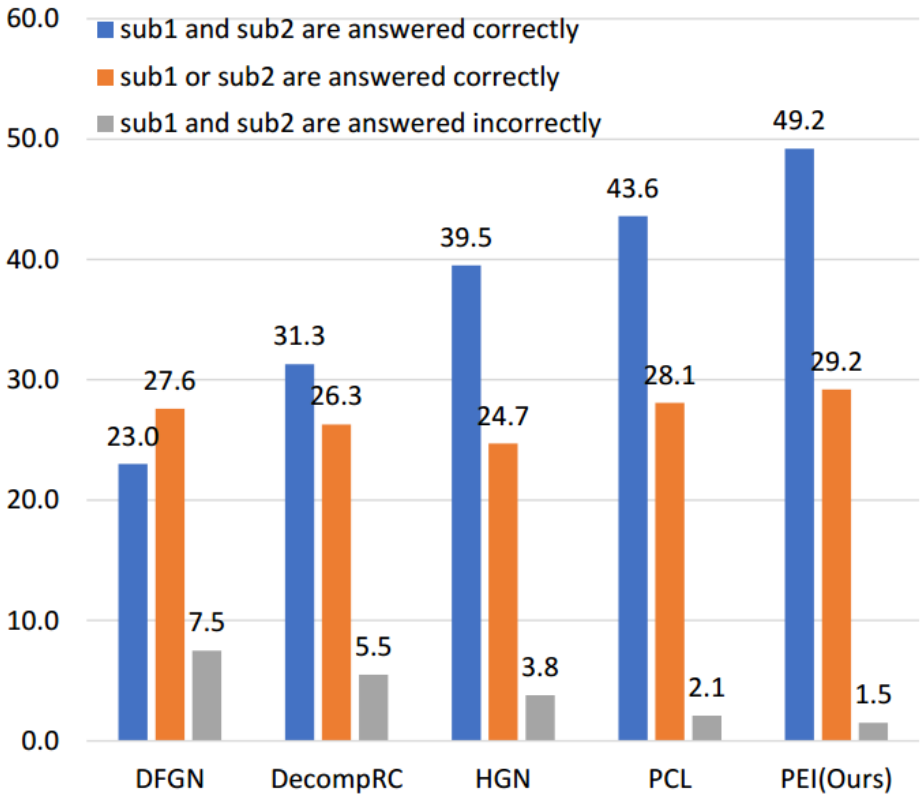
Results: Evaluation of Robustness

Evaluation on Other Multi-hop Datasets

Models	2WikiMultihopQA		MuSiQue	
	EM	F1	EM	F1
iCAP	42.80	47.90	-	-
HGN	38.74	68.69	39.42	65.12
PCL	46.03	73.42	41.28	67.34
PEI (Ours)	**47.32**	**74.56**	**41.97**	**67.85**

unreliable reasoning shortcuts

Evaluation on Sub-question Dataset



Ablation Studies

Effect of Implicit Knowledge

Effect of Type Prompts

Effect of Pre-training on Single-hop

Model	Ans F1	Sup F1	Joint F1
ELECTRA	78.12	88.20	73.50
- Type Prompter	81.14 $\uparrow_{3.02}$	89.37 $\uparrow_{1.17}$	75.68 $\uparrow_{2.18}$
- Pre-trained	78.82 $\uparrow_{0.70}$	88.82 $\uparrow_{0.62}$	74.54 $\uparrow_{1.04}$
- Implicit knowledge	81.22 $\uparrow_{3.10}$	90.25 $\uparrow_{2.05}$	75.80 $\uparrow_{2.30}$
PEI	**85.68** $\uparrow_{7.56}$	**92.11** $\uparrow_{3.91}$	**79.02** $\uparrow_{5.52}$

Ablation Studies

Effect of Foundation PLMs

Model	Ans F1	Sup F1	Joint F1
HGN (RoBERTa)	82.22	88.58	74.37
HGN (ELECTRA)	82.24	88.63	74.51
HGN (ALBERT)	83.46	89.20	75.79
PCL (RoBERTa)	84.33	90.75	77.12
PCL (ELECTRA)	84.42	91.15	77.76
PCL (ALBERT)	85.47	91.28	78.76
PEI (RoBERTa)	85.61	92.02	78.95
PEI (ELECTRA)	85.68	92.11	79.02
PEI (ALBERT)	**86.23**	**92.25**	**79.11**

Conclusion

- **An effective approach based on human reading process**
- **Comparable performance with SOTA on HotpotQA**
- **Ablation studies supports our hypothesis for the PEI model, grounded in the human reading process.**