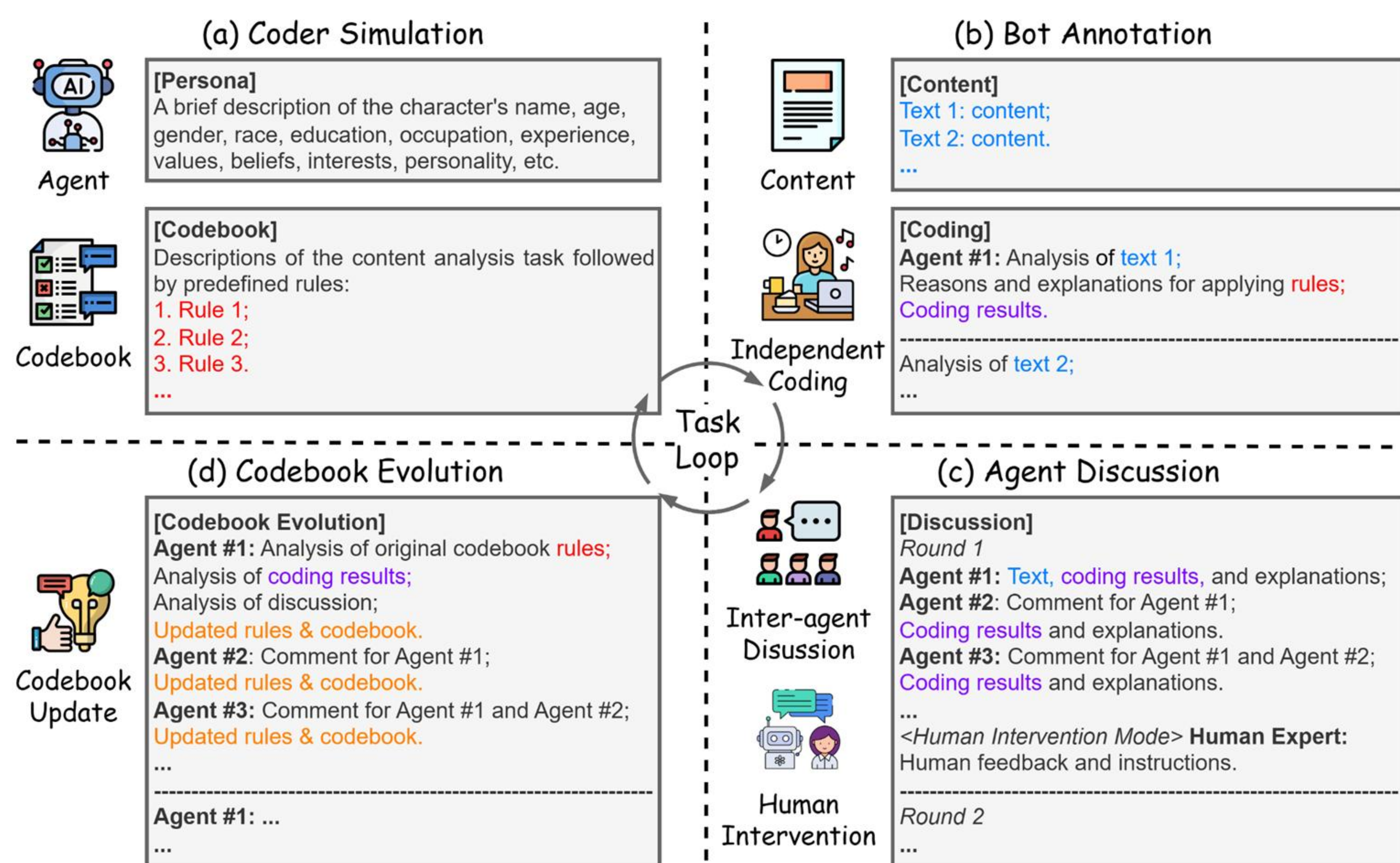# SCALE: Towards Collaborative Content Analysis in Social Science with Large Language Model Agents and Human Intervention

Chengshuai Zhao♠, Zhen Tan♠, Chau-Wai Wong♥, Xinyan Zhao♦, Tianlong Chen♦, Huan Liu♠

♠Arizona State University, ♥NC State University, ♦University of North Carolina at Chapel Hill
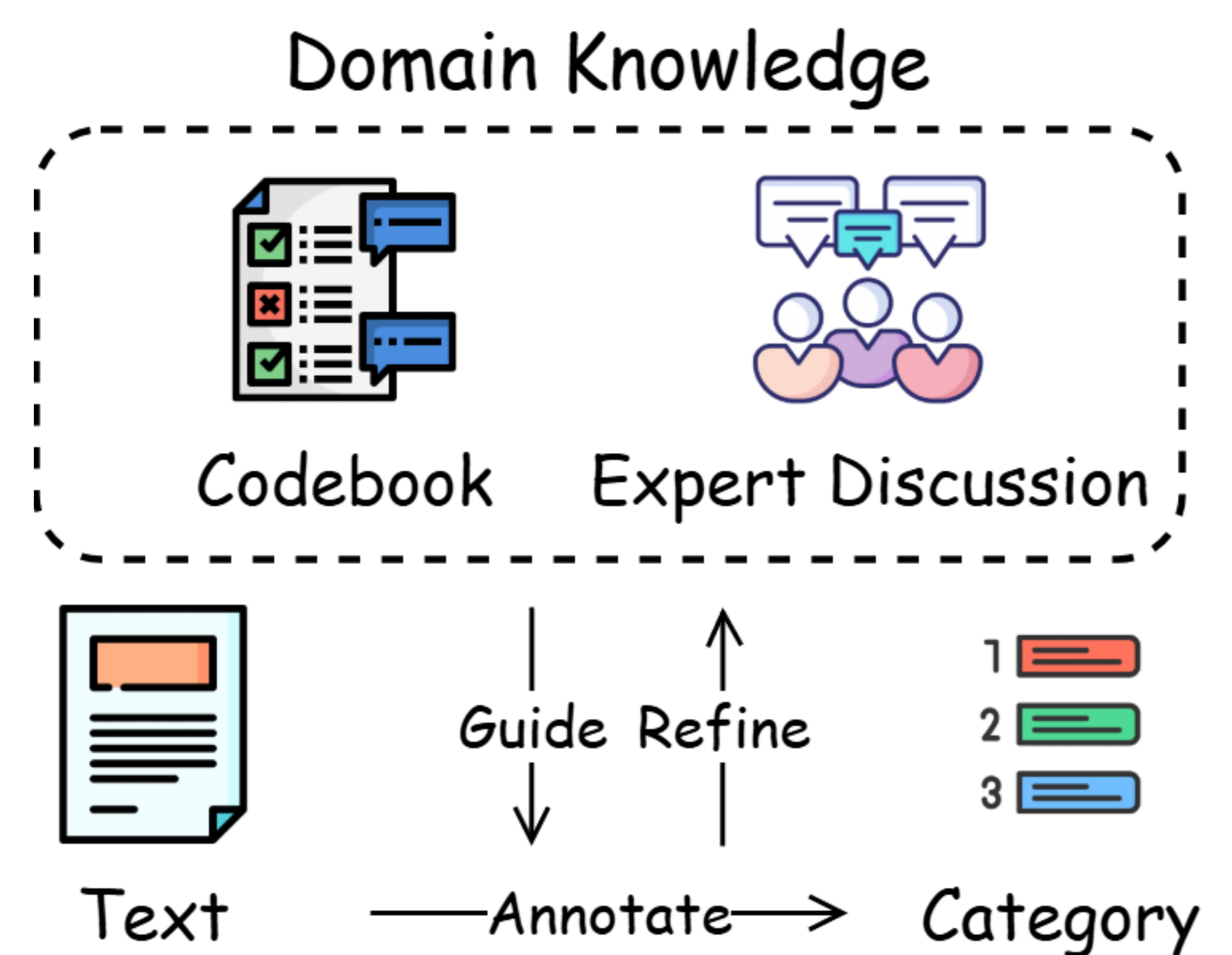
Abstract

Content analysis breaks down complex and unstructured texts into theory-informed numerical categories. Particularly, in social science, this process usually relies on multiple rounds of manual annotation, domain expert discussion, and rule-based refinement. In this paper, we introduce **SCALE**, a novel multi-agent framework that effectively **S**imulates **C**ontent **A**nalysis via **L**arge language model (LLM) ag**E**nts. SCALE imitates key phases of content analysis, including text coding, collaborative discussion, and dynamic codebook evolution, capturing the reflective depth and adaptive discussions of human researchers. Furthermore, by integrating diverse modes of human intervention, SCALE is augmented with expert input to further enhance its performance. Extensive evaluations on real-world datasets demonstrate that SCALE achieves human approximated performance across various complex content analysis tasks, offering an innovative potential for future social science research.
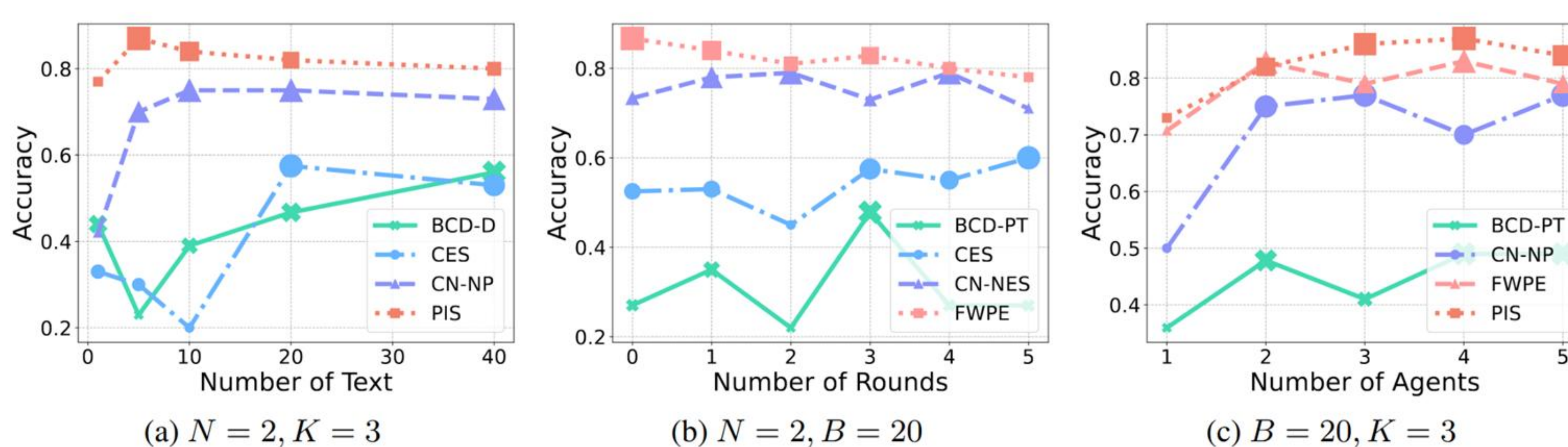
Proposed SCALE framework

(a) Coder Simulation

Agent

Codebook

[Persona]
A brief description of the character's name, age, gender, race, education, occupation, experience, values, beliefs, interests, personality, etc.

[Codebook]
Descriptions of the content analysis task followed by predefined rules:
1. Rule 1;
2. Rule 2;
3. Rule 3.
...

(b) Bot Annotation

Content

Independent Coding

[Content]
Text 1: content;
Text 2: content.
...

[Coding]
Agent #1: Analysis of text 1;
Reasons and explanations for applying rules;
Coding results.

Analysis of text 2;
...

(d) Codebook Evolution

Codebook Update

[Codebook Evolution]
Agent #1: Analysis of original codebook rules;
Analysis of coding results;
Analysis of discussion;
Updated rules & codebook.
Agent #2: Comment for Agent #1;
Updated rules & codebook.
Agent #3: Comment for Agent #1 and Agent #2;
Updated rules & codebook.
...

Agent #1: ...
...

(c) Agent Discussion

Inter-agent Disussion

Human Intervention

[Discussion]
Round 1
Agent #1: Text, coding results, and explanations;
Agent #2: Comment for Agent #1;
Coding results and explanations.
Agent #3: Comment for Agent #1 and Agent #2;
Coding results and explanations.

<Human Intervention Mode> Human Expert:
Human feedback and instructions.

Round 2
...

Task Loop

Content Analysis

Domain Knowledge

Codebook   Expert Discussion

Text → Annotate → Category

Guide ↓ ↑ Refine

1
2
3

Human-Intervened Content Analysis

Intervention Mode		CES	CN-NES	CN-NP	FWPE
No Intervention*		0.63	0.79	0.77	0.91
Targeted	Collabo.	0.73	0.89	0.87	0.95
	Directive	0.73	0.85	0.87	0.95
Extensive	Collabo.	**0.77**	0.89	0.90	**0.96**
	Directive	**0.77**	**0.91**	**0.97**	**0.96**

Automatic Content Analysis

Backbone (w/o intervention)	BCD-PT	BCD-D	CES	CN-NES	CN-NP	FWPE	PIS	Average
GPT-4o	0.51	0.54	0.63	0.79	0.77	0.91	0.88	0.72
GPT-4o w/ COT	0.54	0.61	0.58	0.71	0.52	0.90	0.85	0.67
GPT-4o w/ TOT	**0.57***	**0.63**	0.53	0.73	0.70	0.88	0.87	0.70
GPT-4o w/ self-consistency	0.51	0.57	**0.65**	**0.80**	**0.83**	**0.92**	**0.91**	**0.74**
GPT-4o-mini	0.38	0.47	**0.58**	0.73	0.55	0.79	0.82	0.62
GPT-4o-mini w/ COT	0.19	0.47	0.53	0.72	0.43	0.81	0.71	0.55
GPT-4o-mini w/ TOT	0.35	0.48	**0.58**	**0.83**	0.70	0.84	0.84	0.66
GPT-4o-mini w/ self-consistency	**0.43**	**0.50**	**0.58**	0.79	**0.72**	**0.85**	**0.87**	**0.68**

Discussion Analysis

Parameter sensitivity

(a) $N = 2, K = 3$ (b) $N = 2, B = 20$ (c) $B = 20, K = 3$