

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



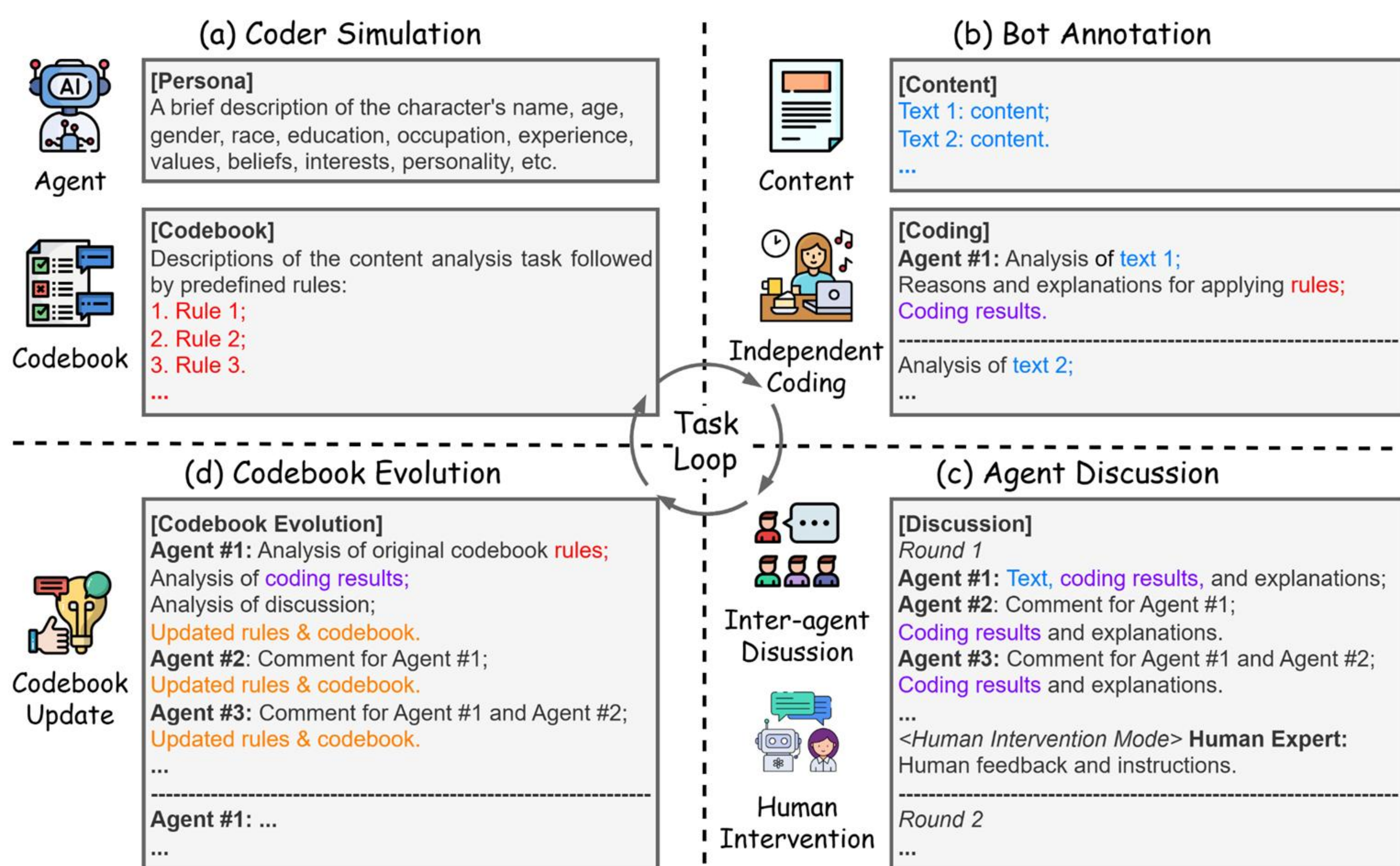
Chengshuai Zhao♠, Zhen Tan♠, Chau-Wai Wong♥, Xinyan Zhao♦, Tianlong Chen♦, Huan Liu♠  
 ♠Arizona State University, ♥NC State University, ♦University of North Carolina at Chapel Hill



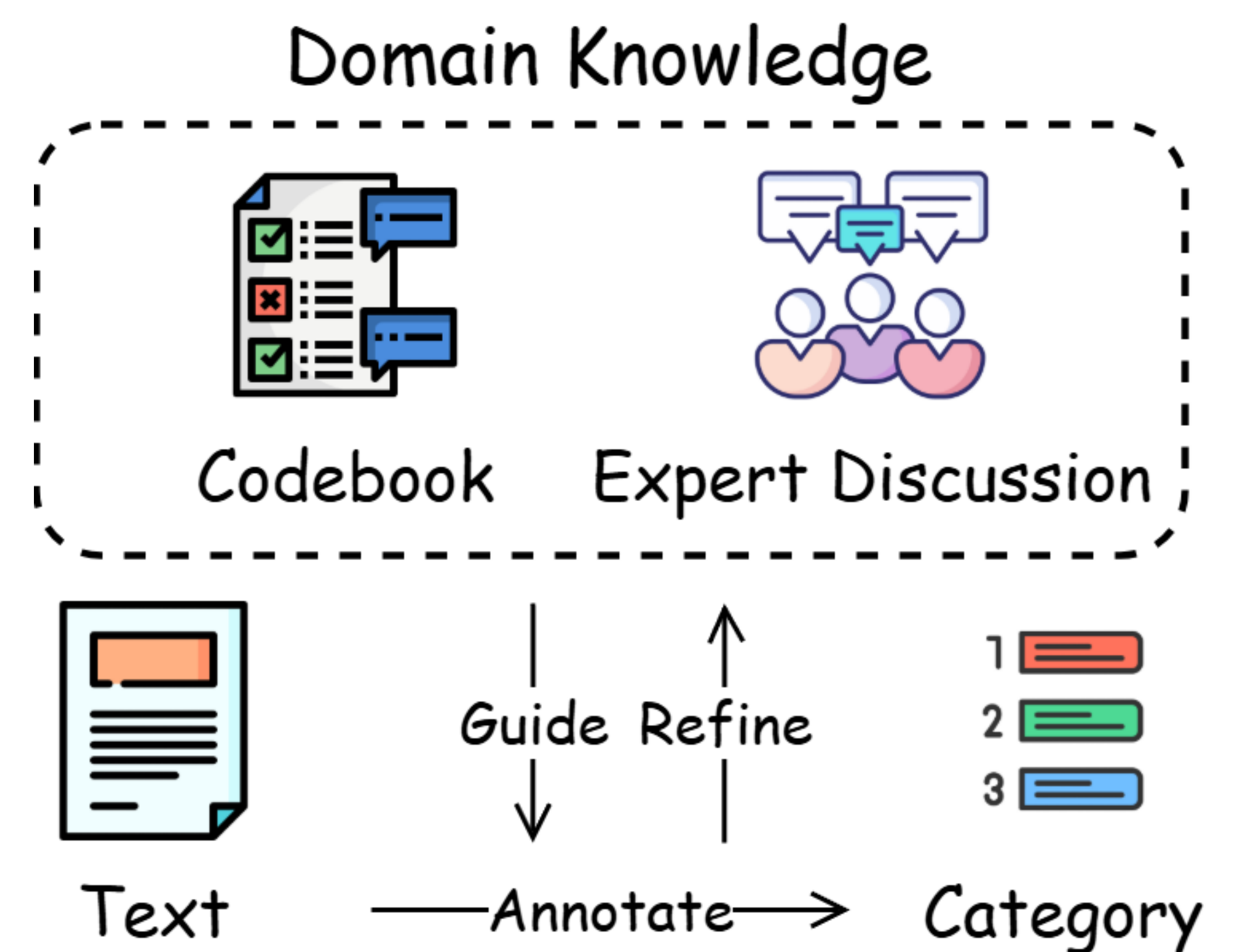
## Abstract

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 Simulates Content Analysis via Large language model (LLM) agEnts. 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



## Content Analysis



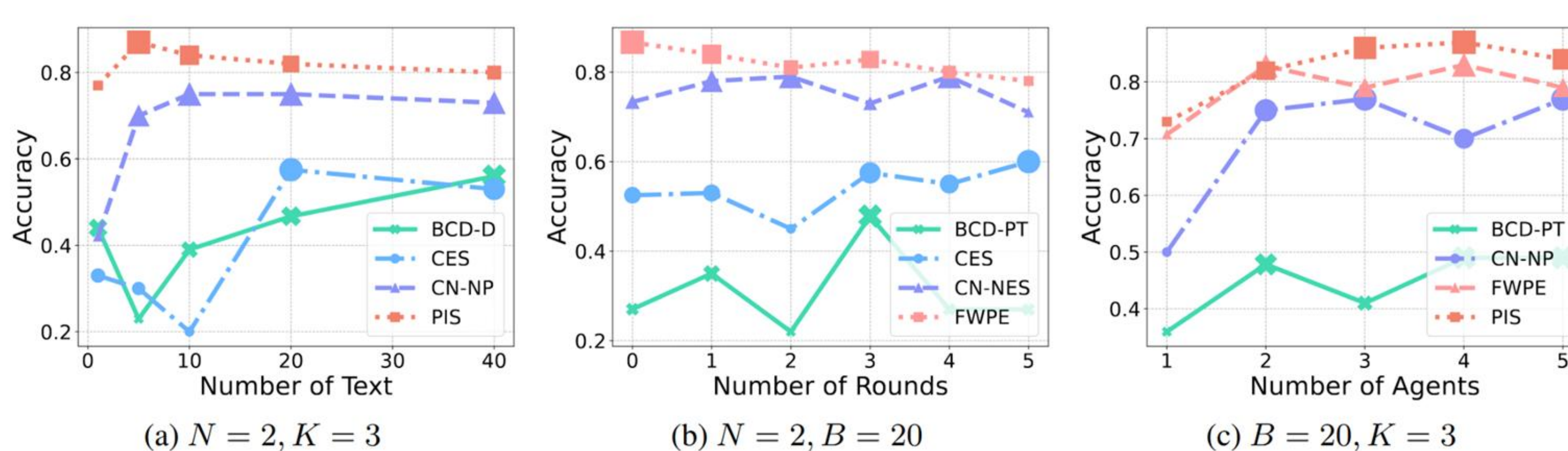
## Human-Intervened Content Analysis

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.	<b>0.77</b>	0.89	0.90	<b>0.96</b>
	Directive	<b>0.77</b>	<b>0.91</b>	<b>0.97</b>	<b>0.96</b>

## 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	<b>0.57*</b>	<b>0.63</b>	0.53	0.73	0.70	0.88	0.87	0.70
GPT-4o w/ self-consistency	0.51	0.57	<b>0.65</b>	<b>0.80</b>	<b>0.83</b>	<b>0.92</b>	<b>0.91</b>	<b>0.74</b>
GPT-4o-mini	0.38	0.47	<b>0.58</b>	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	<b>0.58</b>	<b>0.83</b>	0.70	0.84	0.84	0.66
GPT-4o-mini w/ self-consistency	<b>0.43</b>	<b>0.50</b>	<b>0.58</b>	0.79	<b>0.72</b>	<b>0.85</b>	<b>0.87</b>	<b>0.68</b>

## Parameter sensitivity



## Discussion Analysis

