CharacterGPT: A Persona Reconstruction Framework for Role-Playing Agents

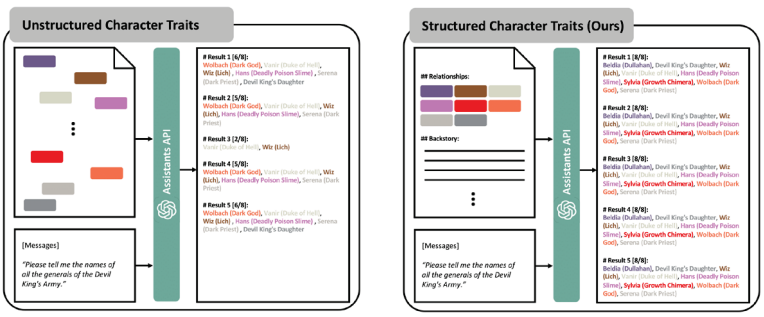
Jeiyoon Park[1], Chanjun Park[2], Heuiseok Lim[2]

naruto@sooplive.com

{bcj1210, limhseok}@korea.ac.kr

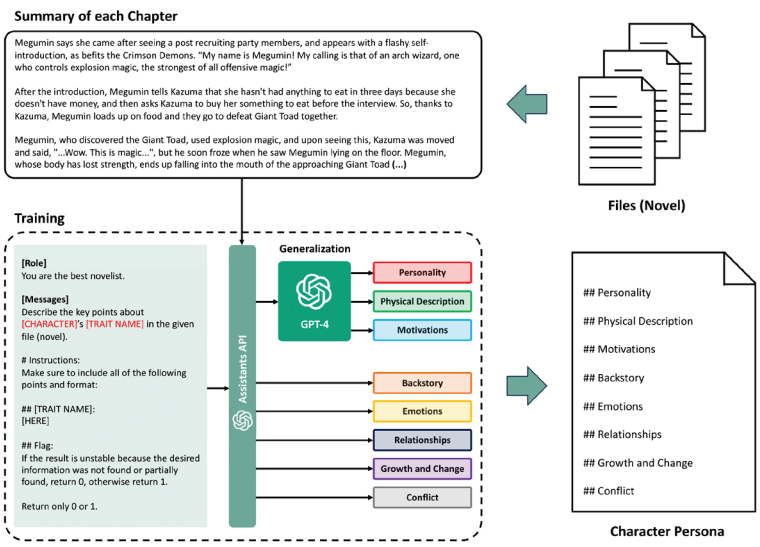
Motivations

- Maintaining consistent character personas remains a challenge due to variability in information extraction, which frequently omits critical elements such as backstory or interpersonal relationships.
- We address the primary research question (RQ) in two key tasks: 1) How to better exploit character persona, and 2) How to encourage characters to use imagination for generating new ideas.



CharacterGPT

- We propose a novel framework called CharacterGPT, which includes a persona rebuilding process called Character Persona Training (CPT).



- Based on character analysis literature, CPT operates by identifying eight essential traits: Internal attributes (Personality, Physical Description, and Motivation) and external attributes (Backstory, Emotions, Relationships, Growth and Change, and Conflict).
- CPT updates the character persona at each epoch by extracting role-specific traits from chapter summaries.
- For Internal attributes (Type A), LLM-based generalization function refines extracted traits, while external attributes (Type B) are simply appended to each persona document.

Experimental Results

- Task for RQ1: Persona Evaluation. For persona evaluation, we compare the personality traits analyzed by human.

Trait	Facets	ChatGPT	ChatGPT+Ours	GPT-4	GPT-4+Ours	Human
OPN	Fantasy	88 (+19)	75 (+6)	75 (+6)	94 (+25)	69
	Aesthetics	69 (+6)	75 (0)	50 (-25)	75 (0)	75
	Feelings	63 (-37)	38 (-62)	69 (-31)	94 (-6)	100
	Actions	50 (-31)	56 (-25)	88 (+7)	94 (+13)	81
	Ideas	63 (-31)	44 (-50)	56 (-38)	81 (-13)	94
	Values liberalism	38 (-6)	44 (0)	38 (-6)	56 (+12)	44
	# Wins	0	3	2	3	-
	Σ|d|	130	143	113	69	-
CON	Competence	50 (-31)	69 (-12)	38 (-43)	69 (-12)	81
	Order	50 (+12)	63 (+25)	44 (+6)	31 (-7)	38
	Dutifulness	50 (-38)	63 (-25)	100 (+12)	94 (+6)	88
	Achievement Striving	63 (-37)	56 (-44)	100 (0)	94 (-6)	100
	Self-Discipline	56 (-19)	50 (-25)	69 (-6)	88 (+13)	75
	Deliberation	50 (+50)	19 (+19)	88 (+88)	56 (+56)	0
	# Wins	0	2	3	2	-
	Σ|d|	187	150	155	100	-
EXT	Warmth	31 (-44)	63 (-12)	88 (+13)	63 (-12)	75
	Gregariousness	38 (-31)	50 (-19)	63 (-6)	50 (-19)	69
	Assertiveness	50 (-31)	63 (-18)	75 (+6)	88 (+7)	81
	Activity	63 (-6)	81 (+12)	63 (-6)	69 (0)	69
	Excitement Seeking	38 (-62)	75 (-25)	100 (0)	88 (-12)	100
	Positive Emotions	50 (-50)	56 (-44)	88 (-12)	100 (0)	100
	# Wins	0	1	3	3	-
	Σ|d|	224	130	43	50	-
AGR	Trust	38 (-43)	50 (-31)	50 (-31)	75 (-6)	81
	Compliance	63 (-12)	50 (-25)	58 (-17)	81 (+6)	75
	Altruism	31 (-38)	63 (-6)	75 (+6)	81 (+12)	69
	Straightforwardness	50 (+12)	38 (0)	100 (+62)	38 (0)	38
	Modesty	63 (+50)	50 (+37)	13 (0)	6 (-7)	13
	Tendermindedness	63 (-25)	44 (-11)	94 (+6)	94 (+6)	88
	# Wins	0	2	3	4	-
	Σ|d|	180	110	122	37	-
NEU	Anxiety	25 (+6)	50 (+31)	13 (-6)	19 (0)	19
	Hostility	63 (-6)	69 (0)	25 (-44)	50 (-19)	69
	Depression	56 (+50)	44 (+38)	75 (+69)	19 (+13)	6
	Self-Consciousness	38 (+38)	50 (+50)	19 (+19)	19 (+19)	0
	Impulsiveness	50 (-31)	50 (-31)	38 (-43)	88 (+7)	81
	Vulnerability	25 (-6)	44 (+13)	38 (+7)	44 (+13)	31
	# Wins	0	1	2	4	-
	Σ|d|	137	163	188	71	-

Table 2: Differences between Megumin's personalities analyzed by humans and LLMs in the BFI test.

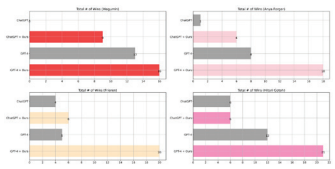


Figure 3: Total sum of # Wins for each character in ChatGPT and GPT-4 settings (Σ # Wins). The larger value, the better.

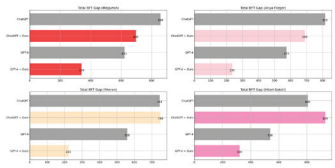


Figure 4: Total sum of $|d|$ for each character in ChatGPT and GPT-4 settings ($\Sigma\Sigma|d|$). The smaller value, the better.

- Task for RQ2: Story Generation. The story generation task is evaluated based on common aspects in generated story assessment.

Model	Grammar	Coherence	Likability	Relevance	Complexity	Creativity
Megumin	3.79	3.82	3.11	**4.21**	2.46	2.86
Megumin + Ours	**4.11**	**4.00**	**3.71**	4.11	**3.46**	**3.29**
Anya	**4.29**	3.82	3.39	3.86	**3.61**	3.68
Anya + Ours	4.25	**4.00**	**3.79**	**4.00**	3.43	**3.89**
Frieren	4.29	3.89	3.50	3.86	3.93	3.79
Frieren + Ours	**4.32**	**3.96**	**3.71**	**4.21**	**4.04**	**3.86**
Hitori	**4.36**	4.04	3.57	**4.18**	3.43	3.50
Hitori + Ours	**4.36**	**4.39**	**3.82**	**4.18**	**3.96**	**3.93**
GPT-4 (avg)	4.18	3.89	3.39	4.03	3.36	3.46
GPT-4 + Ours (avg)	**4.26**	**4.09**	**3.76**	**4.13**	**3.72**	**3.74**

Table 6: Human evaluation of generated stories. The backbone model is the same as GPT-4, and four stories for each setting, a total of 32 stories are generated and evaluated by 7 crowd-workers using a 5-point Likert scale.

Key Advantages

- CharacterGPT minimizes information loss by aligning persona accumulation with narrative progression, and is the first to store and update a protagonist's persona at each epoch, allowing users to engage with characters at specific narrative points.

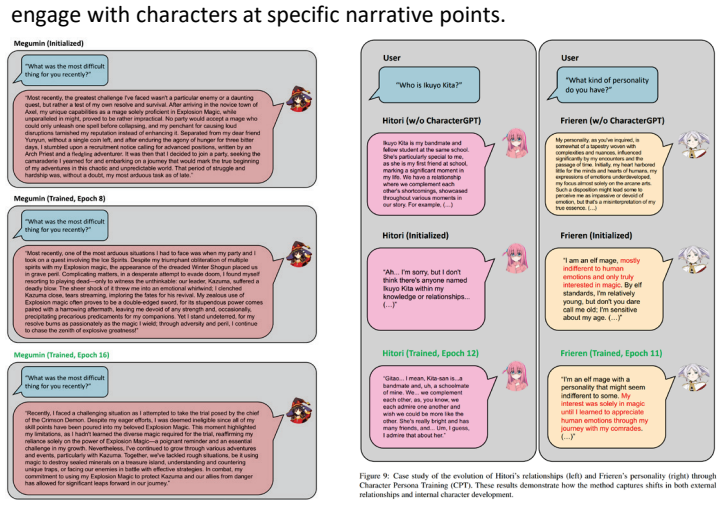


Figure 9: Case study of the evolution of Hitori's relationships (left) and Frieren's personality (right) through Character Persona Training (CPT). These results demonstrate how the method captures shifts in both external relationships and internal character development.

[1]SOOP, [2]Department of Computer Science and Engineering, Korea University

NAACL 2025