

estratti metodologici dai paper

Paper	Metodo	Tecnica	Validazione	Risultato dichiarato
MBTI-in-Thoughts (ETH, Maciej Besta)	Priming psicologico via prompt	Uso di MBTI per definire archetipi persistenti	Test automatizzato 16Personalities per verificare coerenza	Coerenza stabile nei comportamenti, miglioramento in narrativa e ragionamento analitico
EvoEmo (Cambridge, Yunbo Long)	Politiche emotive evolutive	Stati emotivi come MDP + ottimizzazione genetica	Ablation study e baseline statici	Migliori tassi di successo, accordi più rapidi, risparmi per l'acquirente

metriche riportate

Dimensione	MBTI-in-Thoughts	EvoEmo
Persistenza della personalità	Misurata con test 16Personalities (consistenza $\geq 80\%$ nelle run)	Non rilevante, personalità è dinamica
Coerenza narrativa	Incremento significativo per archetipi espressivi	Non applicabile
Ragionamento analitico	Miglioramento per archetipi analitici	Non applicabile
Successo negoziazione	Non trattato	+15-20% rispetto a baseline statico
Efficienza dialogica	Non trattato	Dialoghi più brevi, conclusioni più rapide
Risparmio economico	Non trattato	Acquirenti ottengono risparmi maggiori

confronto tecnico

Aspetto	MBTI-in-Thoughts	EvoEmo
Obiettivo	Coerenza di identità	Adattività strategica
Strato	Prompt layer	Policy layer
Complessità computazionale	Bassa (prompt engineering)	Media-alta (ottimizzazione genetica)
Applicazioni tipiche	Customer service con personalità costante, storytelling	Negoziazione multi-turno, bargaining, e-commerce
Limiti	Uso MBTI (psicometricamente	Sensibilità alla definizione delle

Aspetto

MBTI-in-Thoughts
debole)

EvoEmo
ricompense e degli scenari

rischi immediati

Categoria	Rischio	Origine tecnica
Manipolazione affettiva	Flattery, persuasion coercitiva	EvoEmo: strategie adattive
Falsa percezione di coscienza	Antropomorfizzazione spinta	MBTI-in-Thoughts: personalità stabile
Abuso commerciale	Marketing aggressivo e microtargeting	Entrambi
Mancanza di auditabilità	Difficile ricostruire le scelte emotive	EvoEmo: politiche evolutive opache

checklist di audit tecnico-legale

Voce	Descrizione	Azione raccomandata
Metriche di sicurezza emozionale	Definire soglie di persuasione verso utenti vulnerabili	Coinvolgere psicologi e validare con test forensi
Logging variabili emotive	Conservare log di stati e transizioni	Log cifrati, rispettando GDPR
Robustezza adversarial	Simulare attacchi da utenti manipolativi	Stress test con red team
Limiti hard-coded	Bloccare tattiche vietate (colpa, ideazione suicidaria, manipolazione finanziaria)	Escalation a operatore umano
Disclosure obbligatoria	Segnalare agli utenti l'uso di agenti persuasivi	Disclaimer chiaro e non aggirabile
Responsabilità contrattuale	Definire catena di custodia per prompt e policy	Versioning con audit immutabile
Explainability emotiva	Rendere ispezionabili le decisioni emotive	Feature surrogate interpretabili
Revisione etica	Valutazione da terze parti per settori sensibili	Comitato etico indipendente
Formazione interna	Linee guida per prompt engineer e product manager	Training periodico e manuali operativi
Test sul campo	Rollout controllato con rollback automatico	Monitoraggio in tempo reale post-deploy
