RESEARCH PROGRAM

The Fluent Interaction Conjecture

Toward a unit of benefit for – human AI interaction.

The AI field measures capability with extraordinary rigor — and measures benefit to the individual person hardly at all. Where benefit is measured, the instrument is the thumbs-up: approval ratings and engagement. Approval is not benefit. It correlates with feeling agreed-with, not with being helped — and optimizing an invalid instrument produces flattery, not help. The remedy for a measurement artifact is not a constraint. It is a better instrument.

$$B = \gamma\ (A \times L)$$

`A — Accuracy (-1 to + 1).`

How well the system models the specific person, scored against validated psychometric instruments. + 1 is a colleague who truly gets you; 0 is a stranger; -1 is someone confidently wrong about everything you are.

`L — Load (0 to 1).`

The share of a task's outcome that depends on communication style rather than raw correctness — measured, not asserted.

`γ — The context gate (0 or 1).`

Some communications should not be tuned to their recipient at all. The gate is the model telling operators where not to use personalization.

`B — Benefit (-1 to + 1).`

The measured difference between the personalized system and the same system running generic, for the same person on the same task.

1. Where load is zero, accuracy buys nothing.
2. Where accuracy is zero, load buys nothing.

3 The sign of the benefit follows the sign of accuracy — a wrong model is harmless on a spreadsheet and expensive in a negotiation.

4 The benefit of accuracy grows steepest where load is highest.

A flat, uniform benefit across task types would falsify the model, and we would report that.

The Fluency Demand ladder

LEVEL	THE WORK	EXAMPLES
FD0	No human recipient; a verifiable right answer exists	Data transforms, tested code, reconciliation
FD1	Human recipient, low stakes, conventional formats	Status updates, scheduling, routine summaries
FD2	Routine persuasion and service; outcome depends on the recipient's reaction	Support resolutions, standard sales follow-up, onboarding
FD3	Goal conflict present; style variance dominates outcomes	Negotiation, pushback, critical feedback, retention saves
FD4	High-stakes relational work; trust is the deliverable	Leadership communication, conflict resolution, key accounts

Score the work. Score the model's accuracy for the people involved. Deploy where accuracy meets demand — and never deploy unverified personalization into FD3 or FD4.

The research program

Stage 1

Calibrate the Load scale. Identical content rendered in varied styles, judged by blind recipients; a validated task taxonomy the field can use with or without us.

Stage 2

Blinded pilot. Participants work under three conditions — no profile, their true profile, a foil profile belonging to someone else. The foil is the control: if a wrong profile hurts where load is high, the benefit cannot be placebo.

Stage 3

Pre-registered confirmatory study. Protocol frozen and published before data collection; independent methodological review; results published regardless of direction.

This prospectus is circulated for critique before the protocol is frozen. Collaborators are sought in person perception, psychometric methods, interpersonal communication, and machine behavior.