ACM Symposium on User Interface Software and Technology (UIST) '25

Busan, Korea 30 September 2025

StepWrite: Adaptive Planning for Speech-Driven Text Generation

Hamza El Alaoui

helalaou@cs.cmu.edu

Atieh Taheri

Yi-Hao Peng

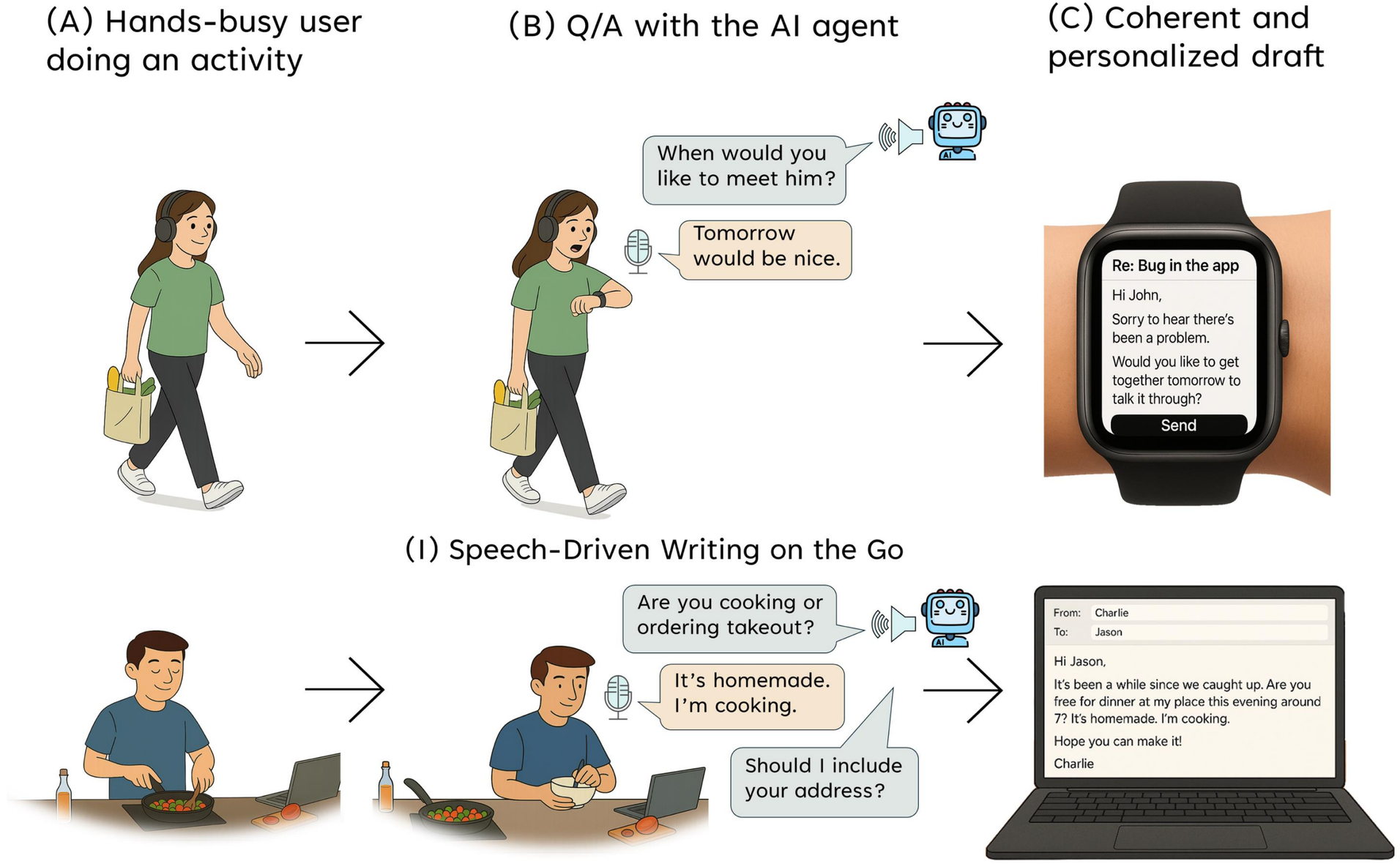
Jeffrey P. Bigham

Carnegie Mellon SCHOOL OF COMPUTER SCIENCE

HCII Human-Computer Interaction Institute

BIG Beneficial Intelligence Group

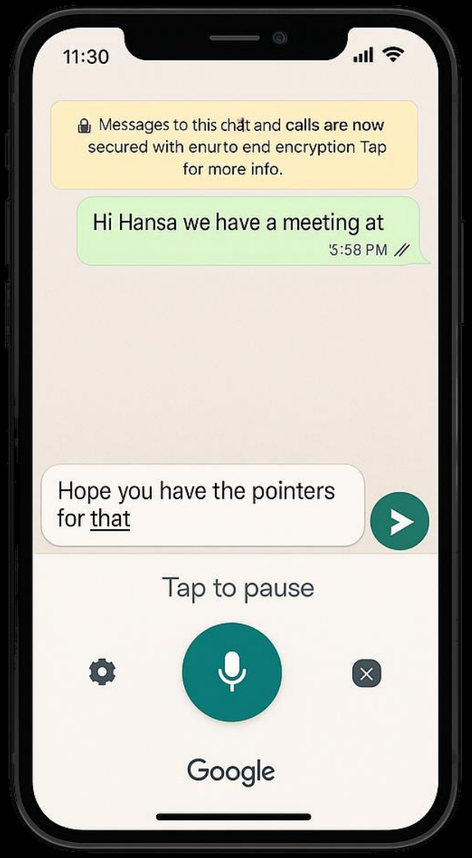
Augmenting your ability to write—anytime, anywhere.



PART ONE

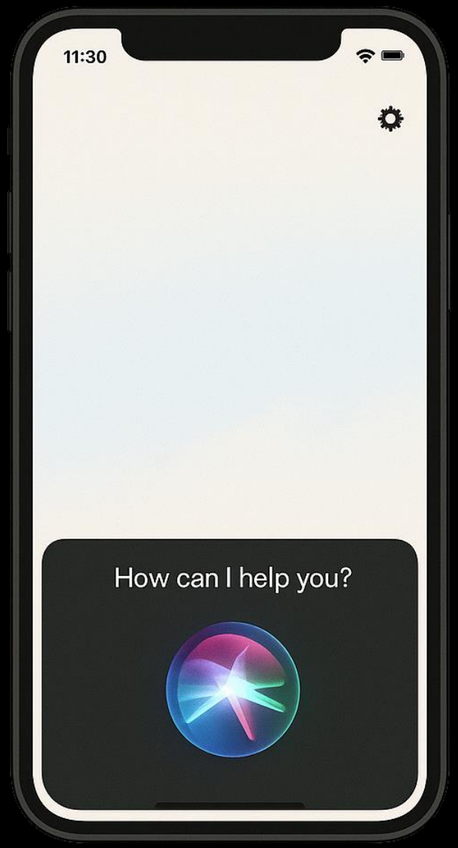
Background

Current Voice Tools



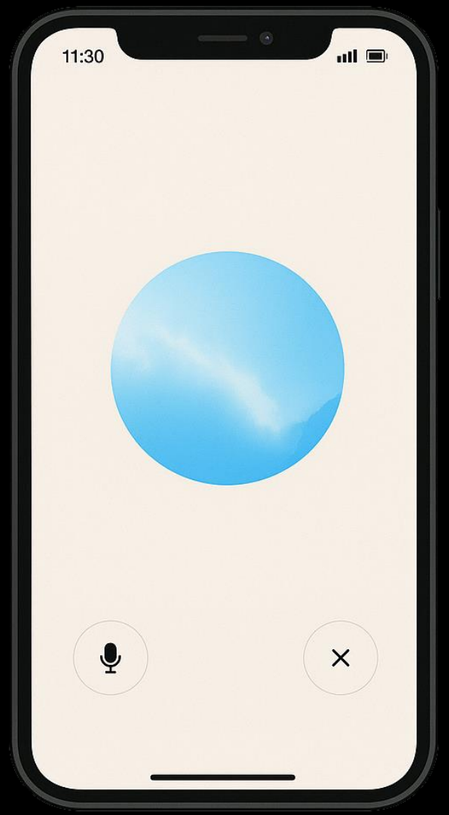
Dictation

linear transcription;
no planning/organization



Voice assistants

command–response;
no memory of goals

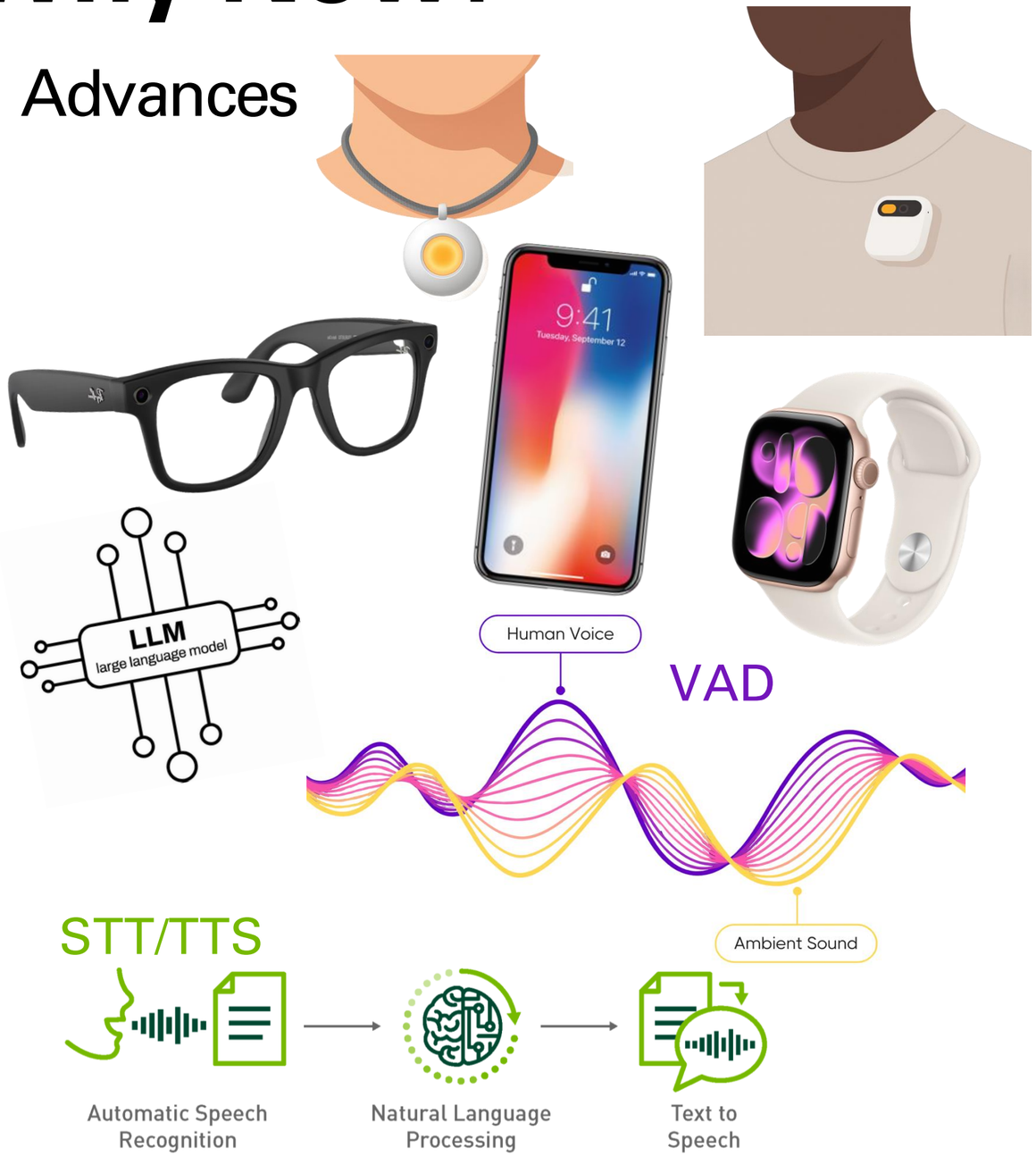


Screen-first AI tools

assume visual attention;
limited scaffolding

Why Now?

Advances



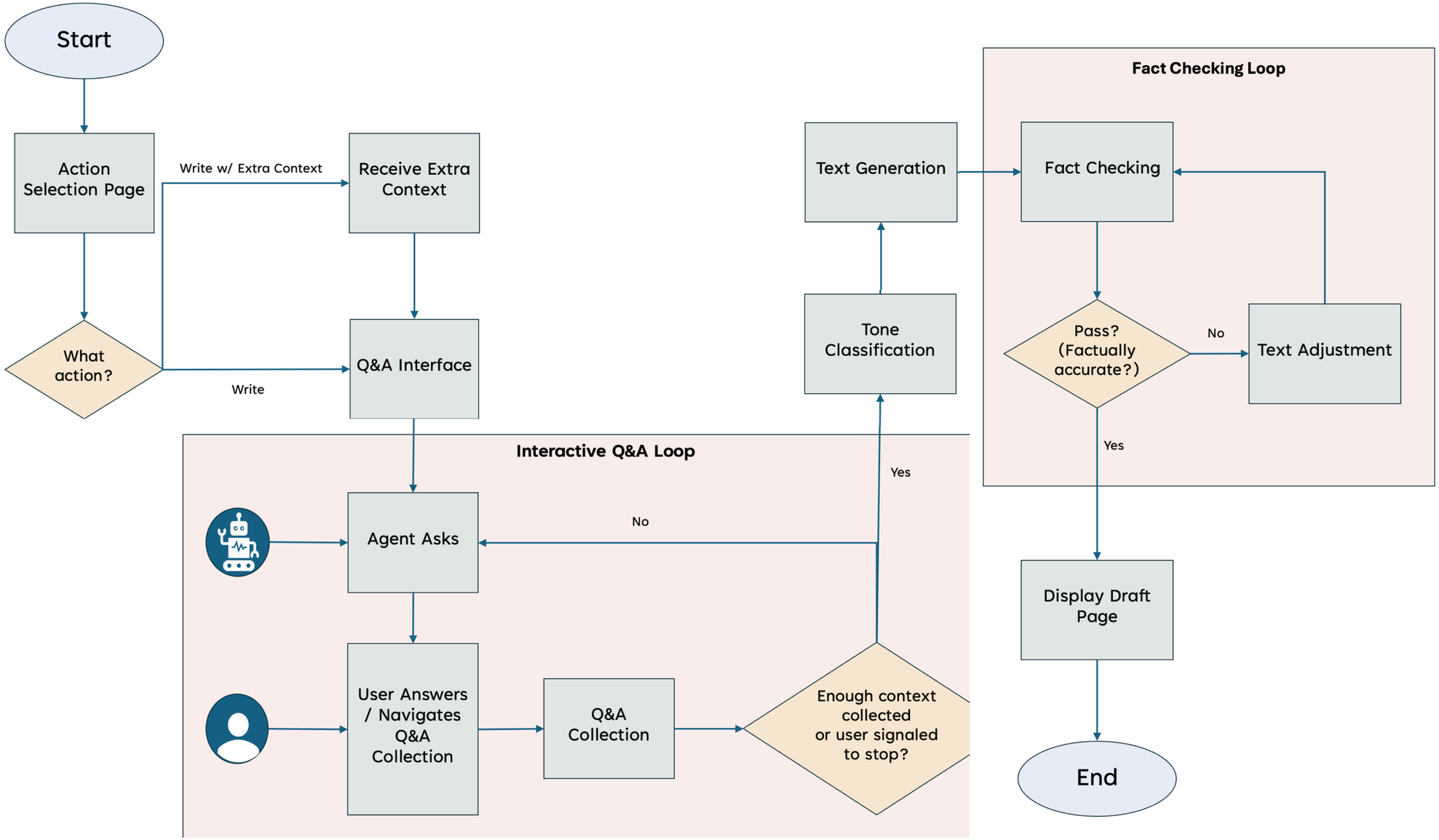
Needs

- Maintain state across turns
- Scaffold planning adaptively
- Support hands-free and eyes-free interaction
- Align tone/intent while preserving authorship

PART TWO

StepWrite

High-level architecture



Demo

* Demo is shown at faster speed for brevity. The full, real-time version is available on the project website.

Voice Commands

Navigation
“next question,” “previous question,” “skip question”

Session Control
“pause writing,” “continue writing,” “finish writing”

Audio Playback
“play that again,” “stop speaking”

Answer Editing
“modify answer”

Interface Switching
“go to editor,” “return to questions”, “show current draft”

* Fuzzy matching; no need for exact phrasing.
  Custom commands possible.
  Future: gestures or sound cues.

PART THREE

Study Design

Setup

Participants.
25 individuals from university community, diverse backgrounds, everyday writing tasks

Tools.
StepWrite, ChatGPT Voice, MS Word Dictation

Tasks.
2 tasks: Write (compose email from scratch) and Reply (respond to provided message)

Contexts.
Stationary (shopping, origami, snacking, Rubik's) and Movement (walking in defined area)

Procedure

Drafting-phase (voice-only).
StepWrite, ChatGPT AVM, MS Word

Revision Phase
Participants edited the draft into something they would personally send

Measures.

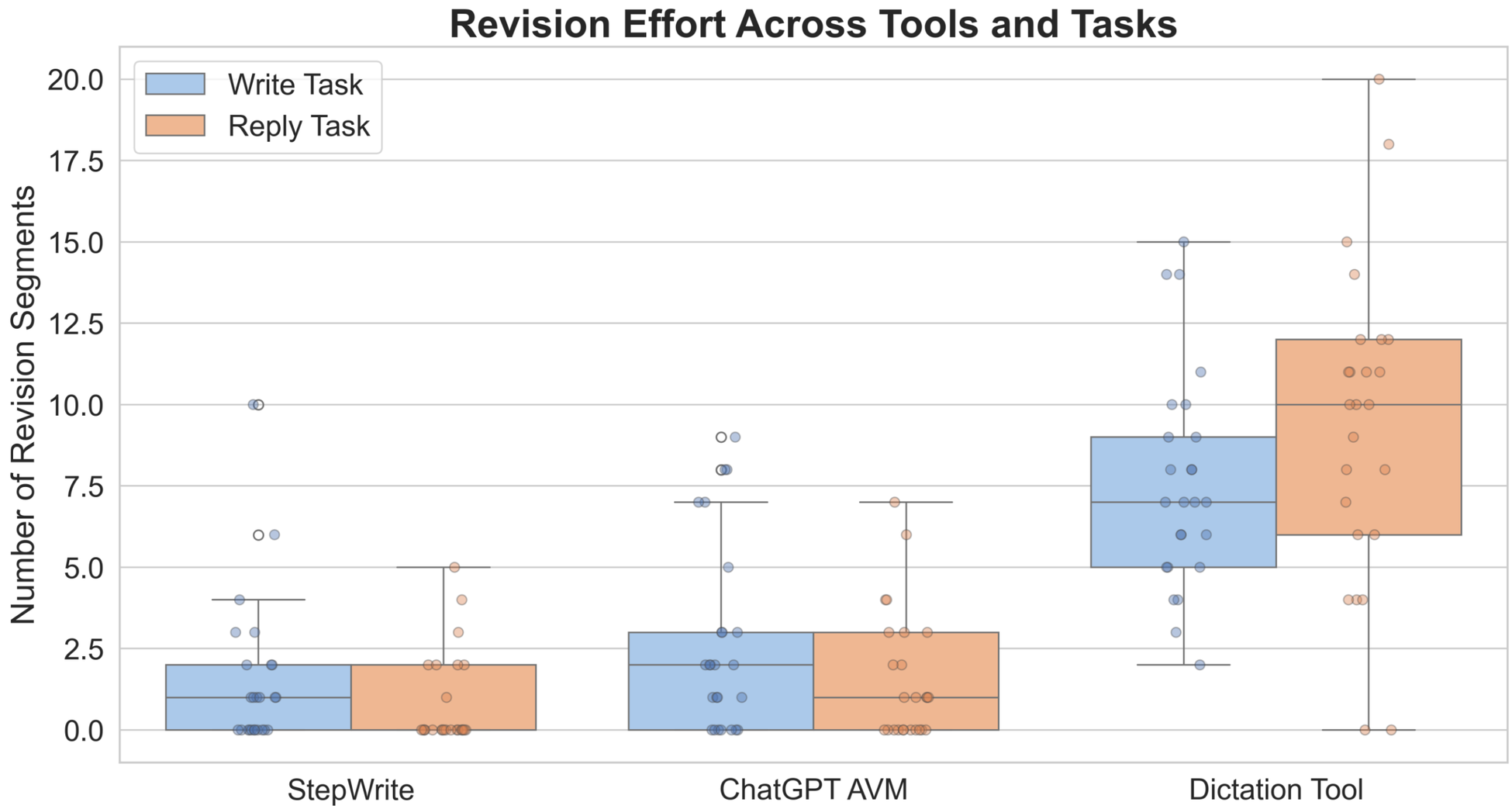
Objective: # of edits, readability scores, sentence structure, lexical/semantic diversity, word count, time

Subjective: workload, usability, emotional engagement, hands-free writing tools assessment (HFWTA)

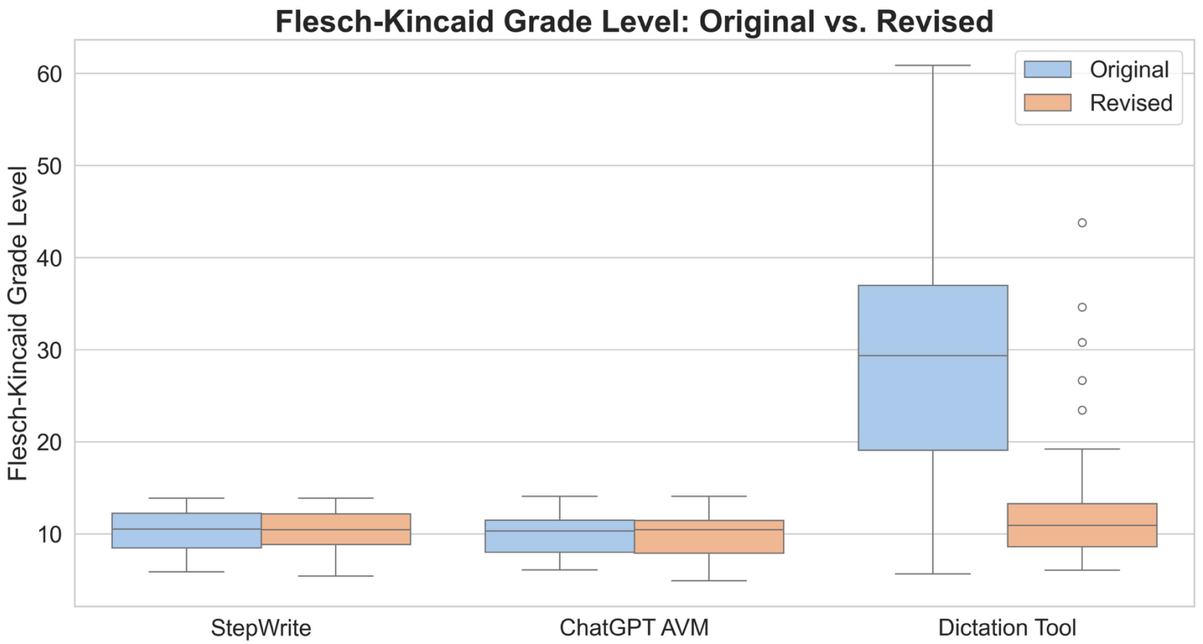
PART FOUR

Results (objective)

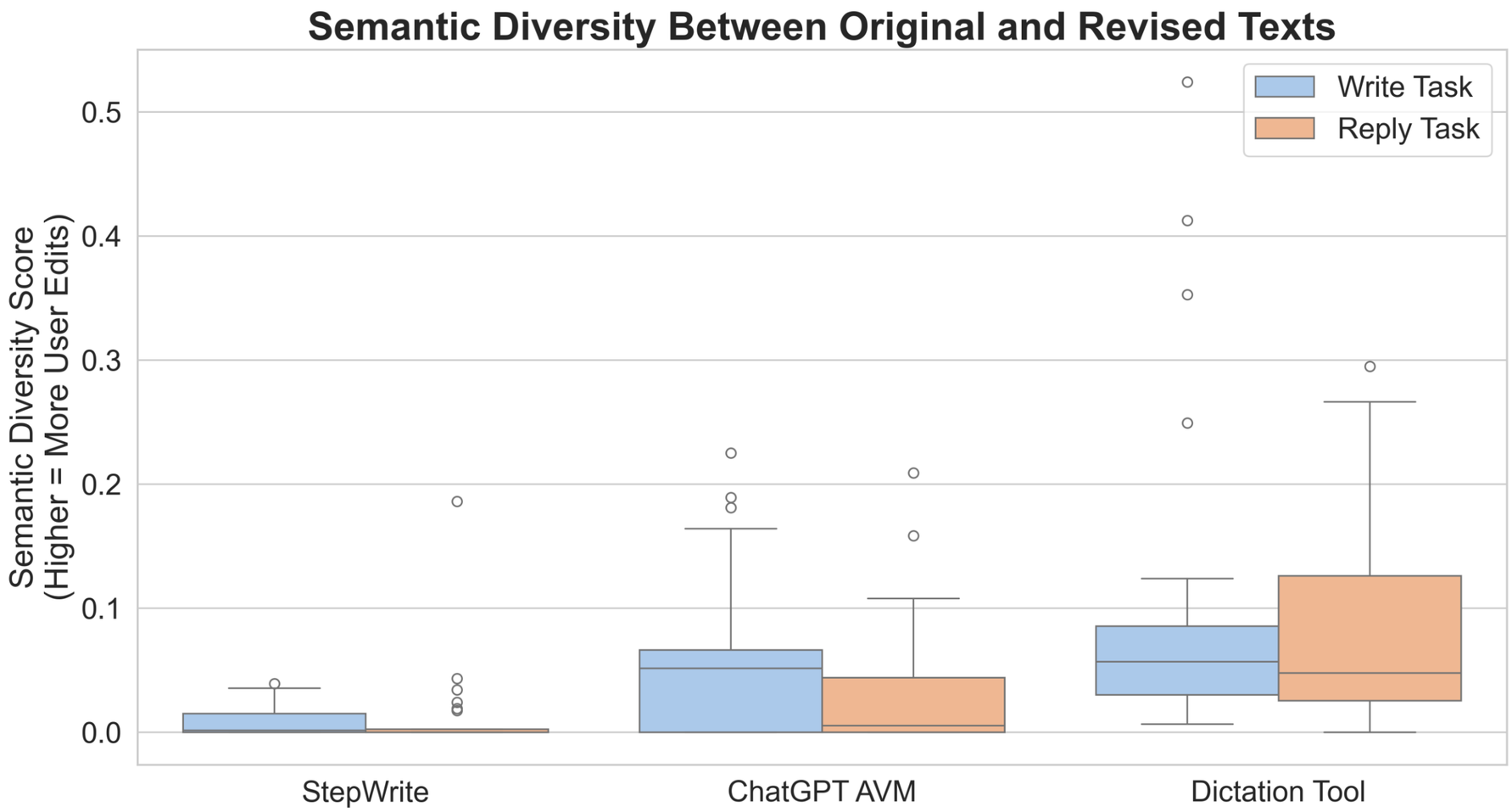
Revision Effort



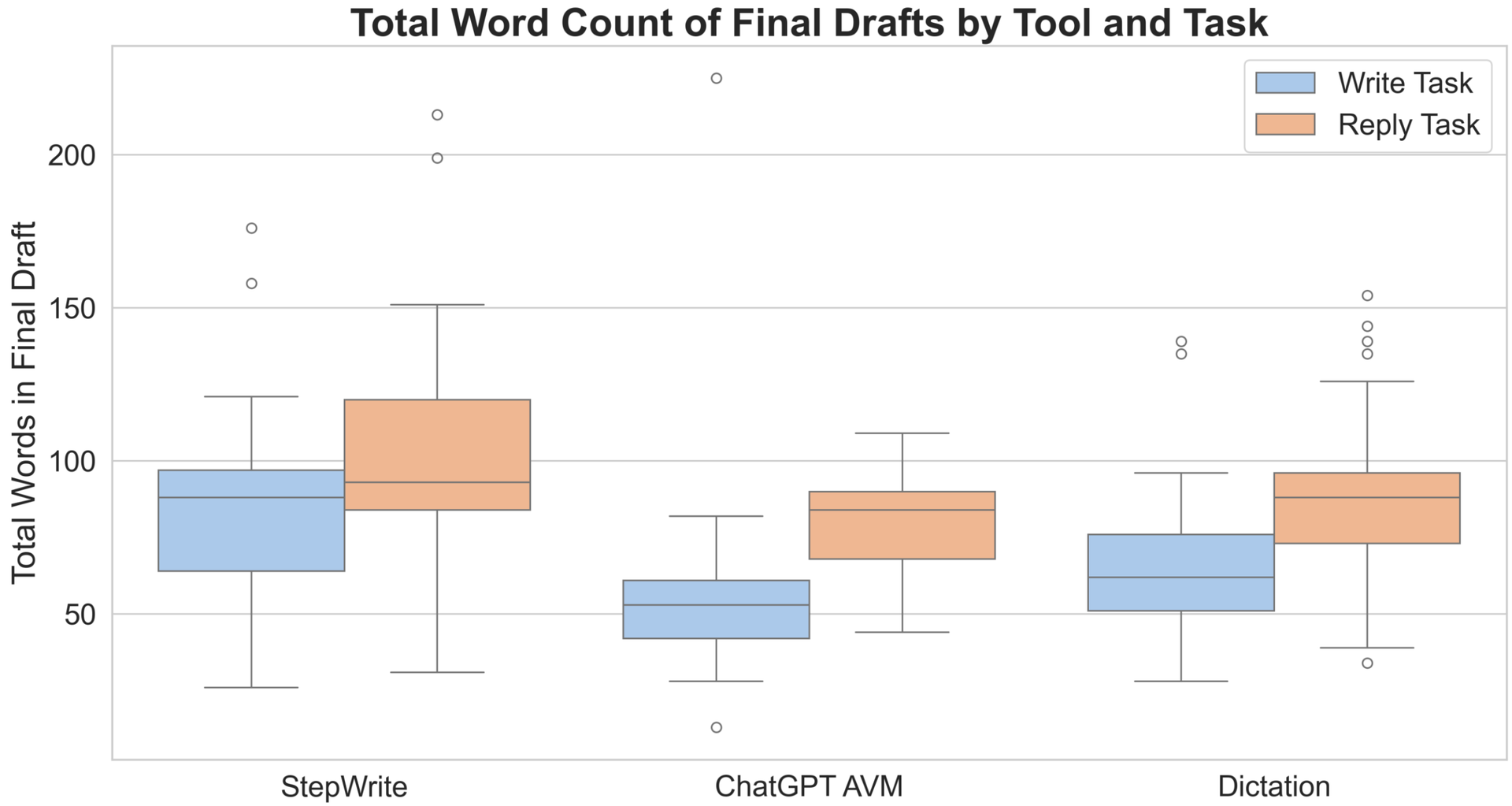
Readability



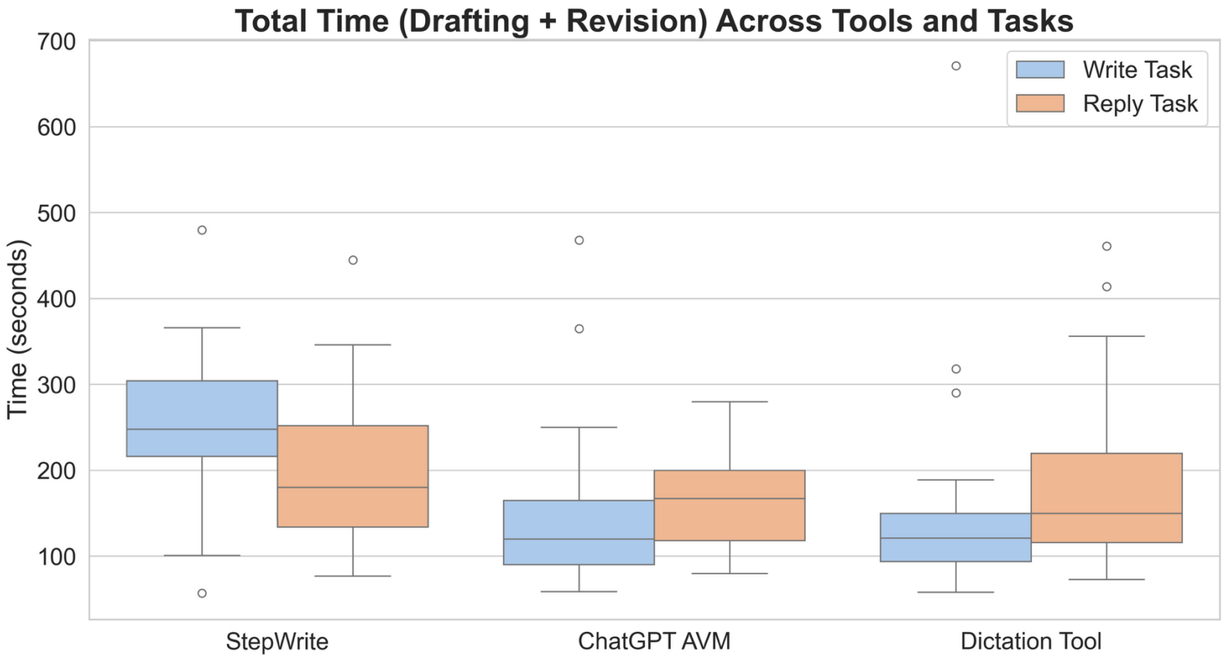
Semantic Diversity

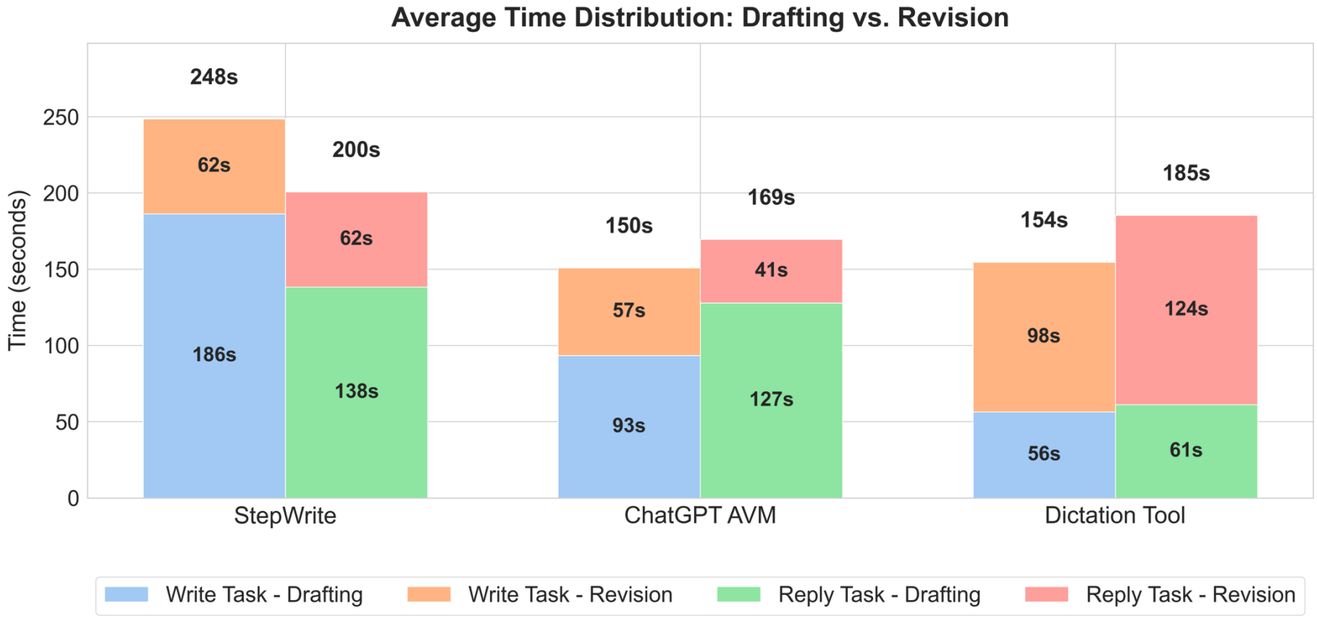


Final Draft Length



Temporal Efficiency





Necessity of Questions

Task	Necessary	Skipped	Unnecessary
Write	74.4%	12.1%	13.6%
Reply	81.8%	11.4%	6.8%

Tone Classification

Overall Accuracy: 0.917

Macro F1: 0.912

Weighted F1: 0.912

Tone	Precision	Recall	F1	Support
Apologetic	1.00	1.00	1.00	22
Encouraging	1.00	1.00	1.00	21
Surprised	0.95	1.00	0.98	21
Cooperative	0.95	1.00	0.98	21
Optimistic	1.00	0.95	0.98	22
Empathetic	1.00	0.95	0.97	20
Concerned	1.00	0.94	0.97	16
Friendly	0.93	1.00	0.96	25
Diplomatic	0.94	0.94	0.94	16
Formal	0.85	0.96	0.90	83
Curious	1.00	0.75	0.86	24
Urgent	0.74	0.95	0.83	21
Informal	0.83	0.79	0.81	19
Assertive	1.00	0.42	0.59	19

PART FIVE

Results (subjective)

User Perceptions

Workload

StepWrite least effortful; Dictation most demanding.

Usability

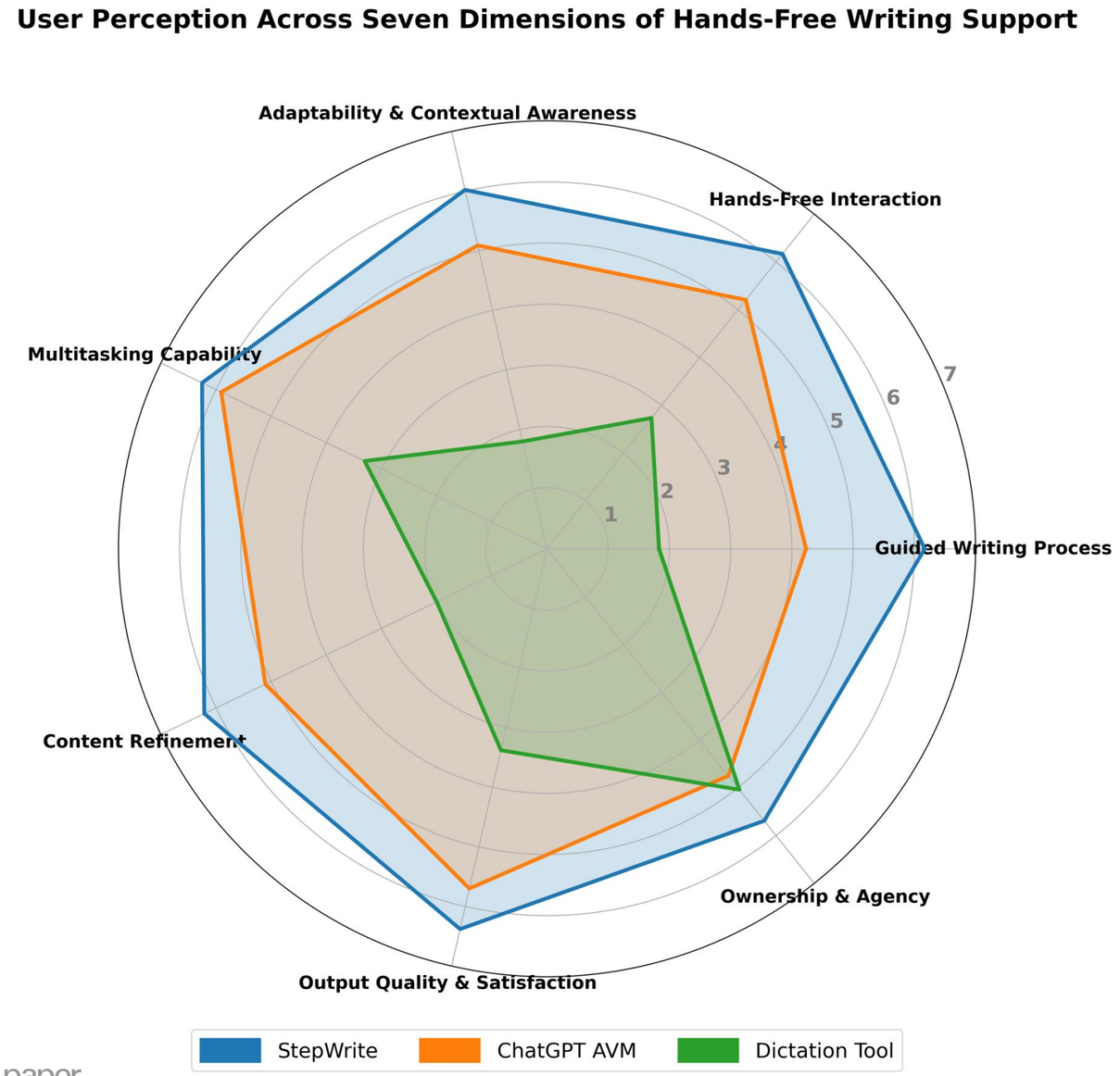
StepWrite & ChatGPT rated “Excellent”; Dictation below average.

Experience

Experience: StepWrite most engaging; ChatGPT reduced stress; Dictation frustrating.

* Detailed subjective evaluation can be found in the paper.

Hands-Free Writing Tools Assessment



* This instrument is released with the paper.

Where People Want StepWrite

Scenario	Activities	N
Email / messaging	drafting or replying to emails, DMs	21
Gym / workouts	treadmill, weight-lifting, gym bike	9
Hands-full multitasking	carrying items, “hands were full”	8
Idea capture / notes	brainstorming, jotting rough text	7
Driving / commuting	driving a car, riding a bus	6
Walking / on-the-go	campus walks, errands	5
Cooking / food prep	meal prep, stirring pots	5
Urgent replies	replying quickly, deadlines	4
Academic writing	class assignments, school papers	4
Lab / technical work	soldering, molten-steel experiments	3
Running (outdoors)	jogging, road running	2
Accessibility needs	wheelchair use, fatigued hands	2
Quiet desk environment	low-noise desk work	2
Weather conditions	winter weather, gloved hands	2
Planning / to-do lists	vacation planning, daily task lists	1

PART SIX

Findings and Future Directions

Key Findings and Design Insights

StepWrite → Cleanest drafts, lowest workload, most engaging

Use cases → Email/text, multitasking, accessibility, safety-critical work

Design insights →
- Adaptive scaffolding > static templates
- Transparency builds trust with users
- Flexibility and micro-edits reduce fatigue

Broader Implications and Future Vision

Structured prompts reduce edits → (–77% vs Dictation, –40% vs ChatGPT)

Accessibility impact → Voice-only workflow aids dexterity and memory challenges

Future vision → A new input modality, especially for wearables and mobile

Open source → StepWrite released for community use! 🤗

Special thanks to

Funding and support

Compute resources

Takeaways.

- StepWrite scaffolds long-form writing with targeted Q&A

- Q&A scaffolding improves writing in multitasking contexts, a promising input modality on-the-go writing/wearables

- Outperforms both dictation and ChatGPT

- Released system & evaluation instruments as open-source

Q&A

Project Website

Contact
helalaou@cs.cmu.edu

On the market for opportunities

Atieh Taheri, Ph.D.
atiehtaheri.com
Seeking academic/industry positions
Accessibility, AR/VR, AI/ML

Hamza El Alaoui
hamzaelalaoui.com
Seeking a summer internship
Human augmentation, AI/ML, multimodal systems, robotics

Link to our portfolios

Beneficial Intelligence Group

Thank You

Beneficial
Intelligence
Group