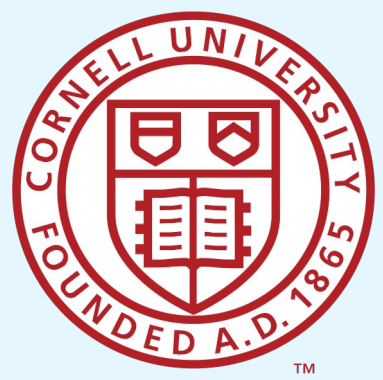
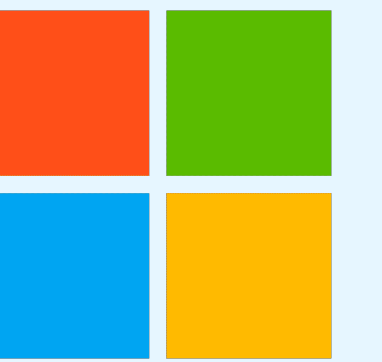




Adopt $\neq$ Adapt:

Longitudinal Analyses of LLM Conversations in the Wild

Rebecca M. M. Hicke and Kiran Tomlinson



Individual users **do not** get better at using LLMs over time

Datasets

Bing Copilot

Apr. 1, 2024 – Sept. 30, 2024

By population

175,650 796,838

Sampled by calendar day

By user

11,905 812,650

Sampled by # days active

WildChat 4.8M

Mar. 9, 2023 – July 31, 2025

1,830,631

users matched by hashed IP

2,522,330

API-like usage removed by date cut off

Features

Syntactic

user messages / conversation

avg. sentence length / conversation

conversations / day

Semantic

User intent

user's purpose for conversing with chatbot

9 categories (e.g. summarization)

Conversational domain

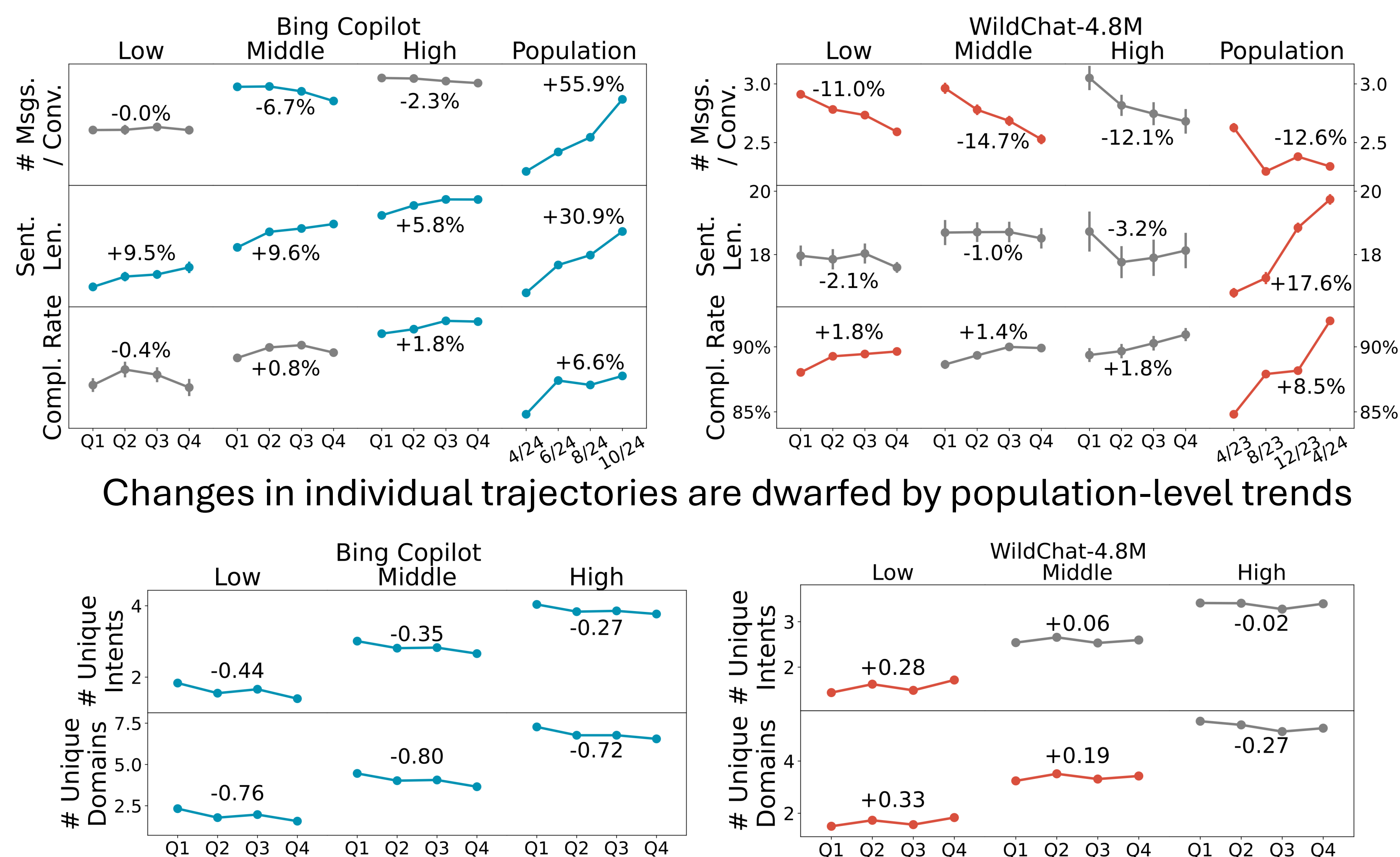
general subject area of conversation

30 categories (e.g. law and politics)

Task completion

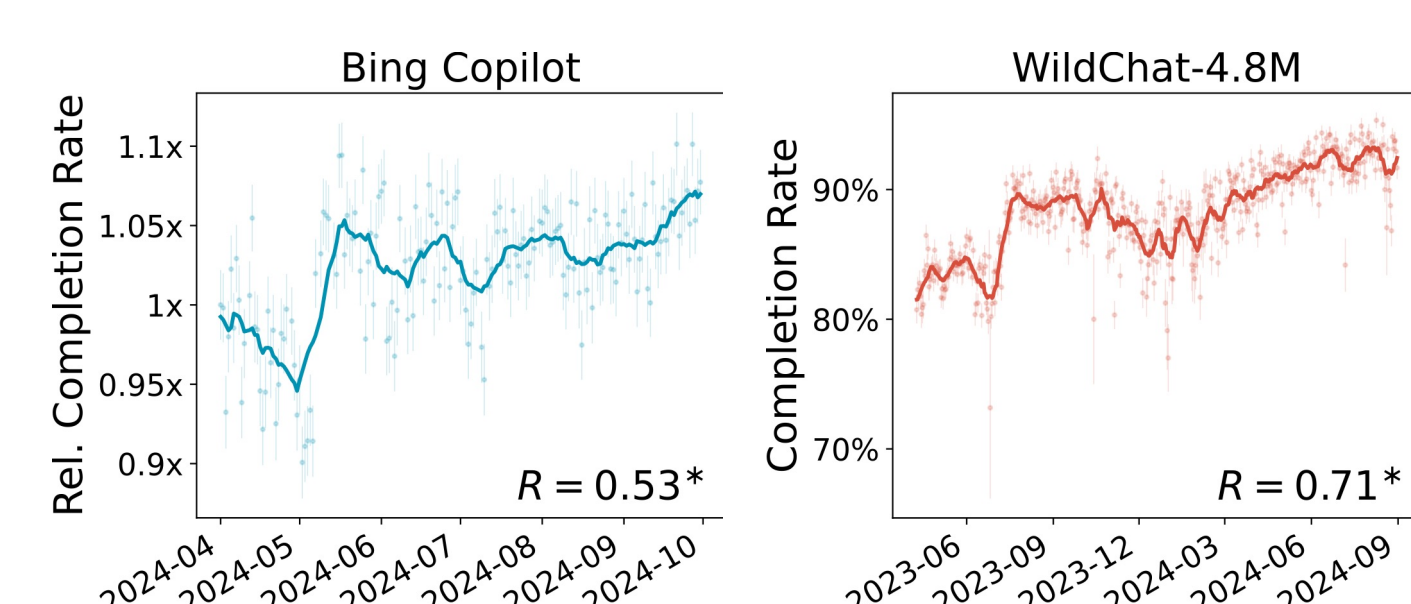
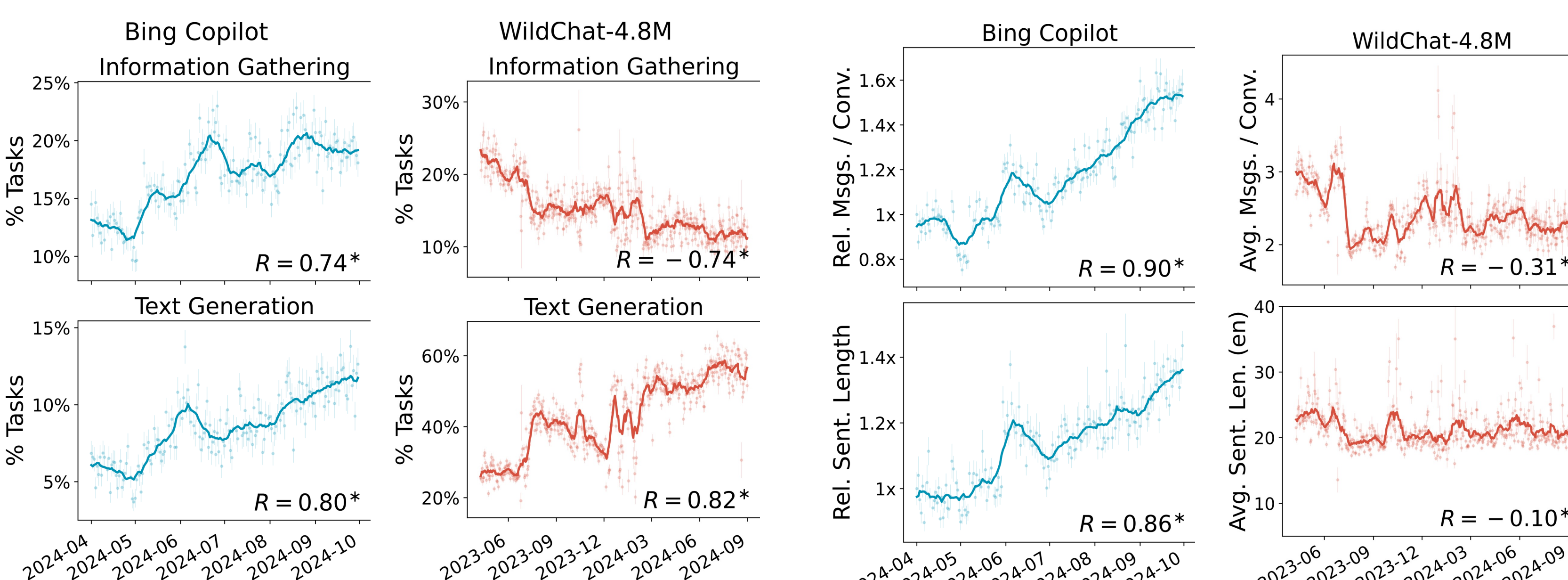
user's primary task is completed by end of conversation

Over LLM use, user behavior changes only slightly

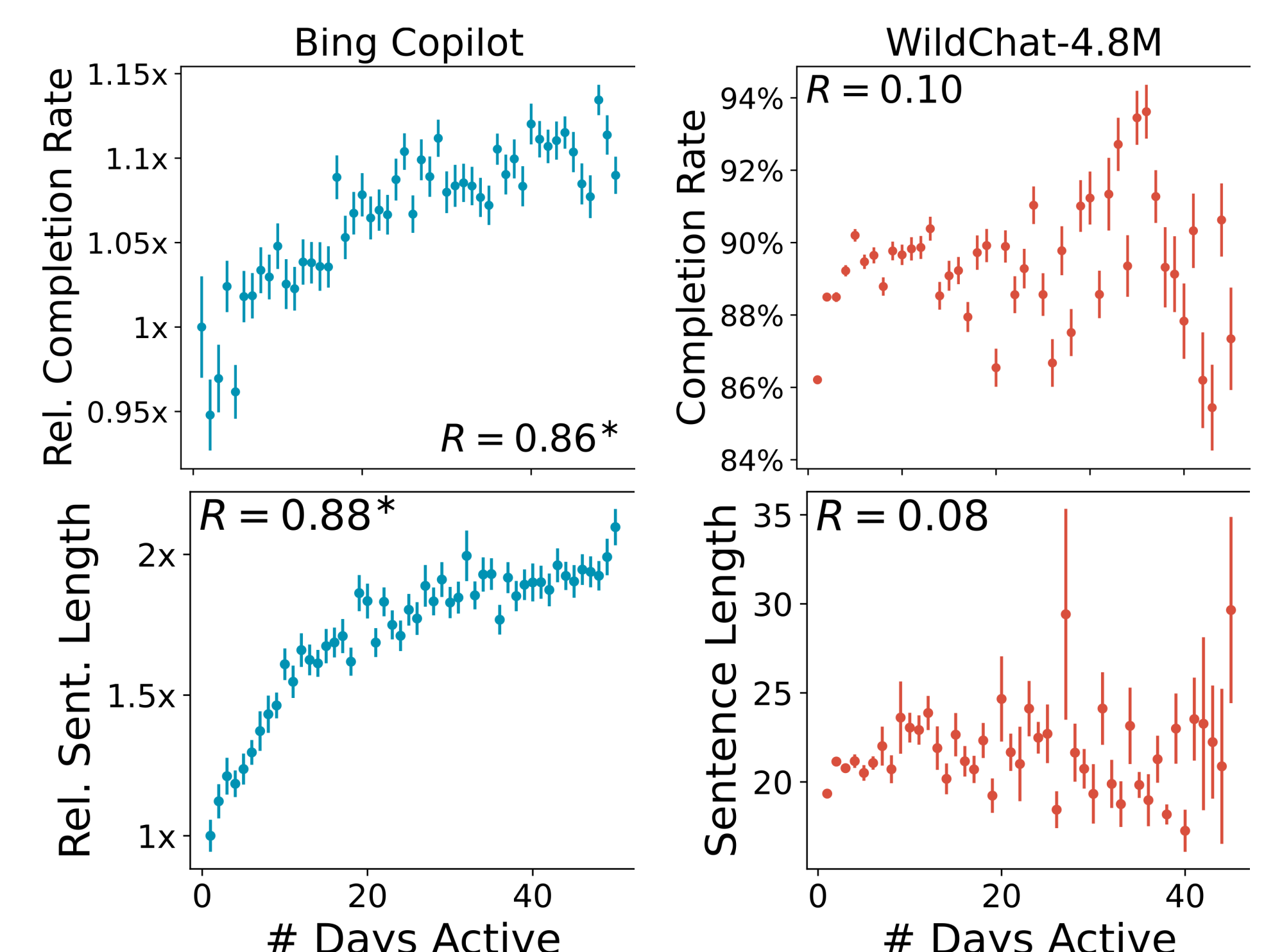


Copilot (but not WildChat) users shift slightly from exploration to exploitation

... and population-level behavior does evolve



... but more active users behave very differently

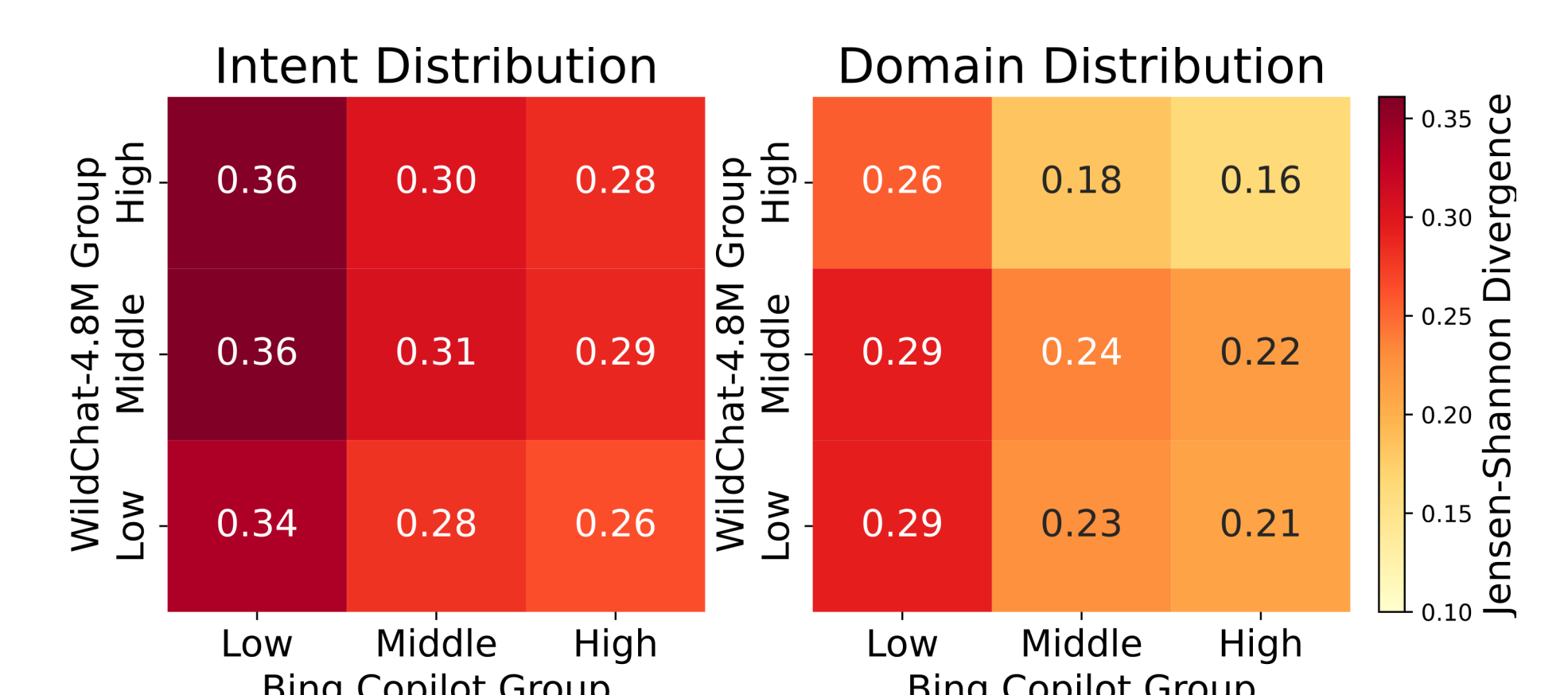


Intent	Bing Copilot			WildChat-4.8M		
	Middle (pp)	High (pp)	Low (%)	Middle (pp)	High (pp)	Low (%)
Summarization	+15.7	+19.3	91.8	+1.2	+3.2	91.8
Translation or Conversion	+7.7	+15.1	89.8	+2.9	+3.0	89.8
Open-Ended Discovery	+6.2	+12.8	70.0	-2.8	-0.0	70.0
Analysis	+2.8	+7.9	86.2	-0.3	+0.9	86.2
Information Gathering	+5.9	+7.5	86.1	+0.4	+1.7	86.1
Text Generation	+2.1	+4.6	94.4	-1.8	-2.2	94.4
Information Lookup	+2.4	+3.6	85.8	+1.2	+1.4	85.8
Image Creation	+0.0	+0.3	83.4	+3.6	+0.5	83.4
Website Navigation	-1.2	-4.8	62.4	+6.6	+15.2	62.4

More active users even have higher completion rates by intent, especially Copilot users

Domain	Bing Copilot			WildChat-4.8M		
	Low (%)	Middle (pp)	High (pp)	Low (%)	Middle (pp)	High (pp)
Programming and Scripting	5.1	+2.6	+3.8	14.5	+1.0	-1.4
Creative Writing and Editing	3.5	+2.0	+2.9	21.6	-2.6	-6.7
Professional Writing and Editing	1.0	+0.7	+1.4	1.0	+0.3	+1.7
Entertainment	4.8	-1.6	-0.7	2.8	+0.5	-0.6
Shopping and eCommerce	1.7	-0.6	-0.8	0.6	+0.5	+0.7
Small Talk and Chatbot	2.4	-0.4	-0.9	2.4	+0.4	-0.2
Travel and Tourism	3.2	-1.7	-2.0	0.6	+0.0	+0.0

More active Copilot users converse more about professional topics and less about casual topics



Check out the paper for full details!