

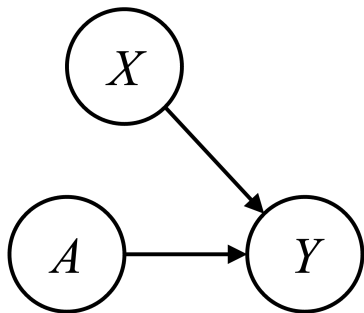
# Generator-Mediated Bandits: Thompson Sampling for GenAI-Powered Adaptive Interventions

Marc Brooks\*, Gabriel Durham\*, Kihyuk Hong\*, Ambuj Tewari

*\*Equal contribution, listed alphabetically*

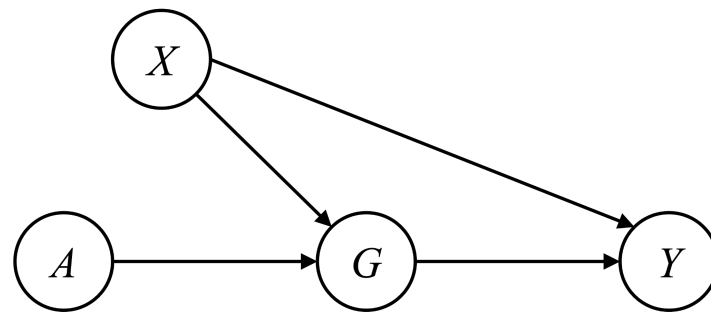
# Generator-Mediated Bandit (GAMBIT)

## Standard Bandit Framework



- Agent observes context ( $X$ ) and selects action ( $A$ ).  $A$  and  $X$  influence reward  $Y$
- Goal: Learn  $Y|A, X$  to select actions that maximize expected reward

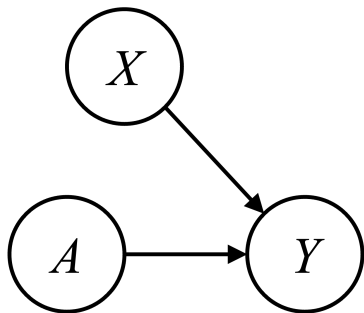
## GAMBIT



- Agent observes context ( $X$ ) and selects generator query ( $A$ ). Generator response ( $G$ ) and  $X$  influence reward  $Y$
- Goal: Learn  $Y|G, X$  and  $G|A, X$  to select actions (generator queries) that maximize expected reward
- Agent cannot directly control  $G$ , only  $A$

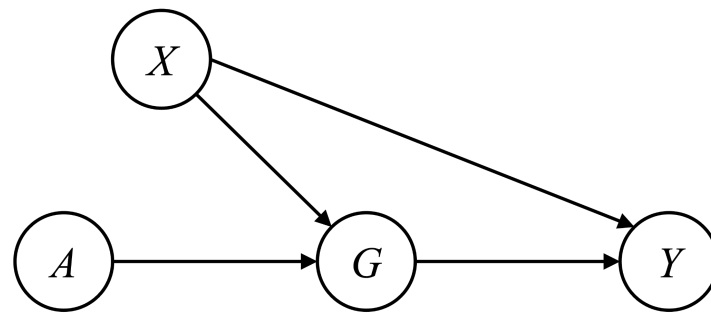
# Motivation - Text-Based mHealth Apps

## mHealth App with Fixed Message Library



- Observe user context ( $X$ )
- Send message to user ( $A$ )
- Observe proximal user health outcome ( $Y$ )

## “LLM-Powered” mHealth App

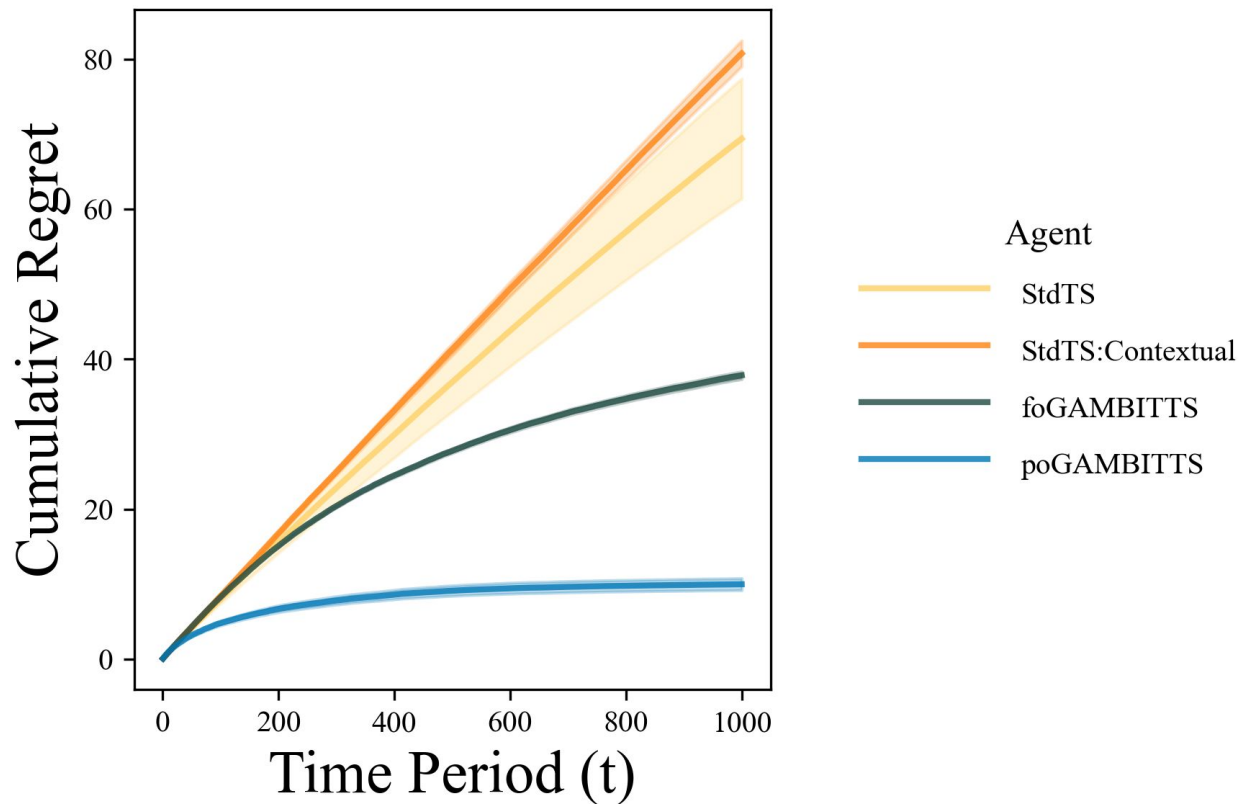


- Observe user context ( $X$ )
- Send query ( $A$ ), along with  $X$ , to LLM
- Observe LLM response ( $G$ ) and send to user
- Observe proximal user health outcome ( $Y$ )

# Method: Generator-Mediated Bandit–Thompson Sampling (GAMBITTS)

- Thompson sampling-based approach, decomposes  $Y|A,X$  to:
  - $G|A,X$  (treatment model), and
  - $Y|G,X$  (reward model)
- Two broad approaches:
  - Fully online GAMBITTS (foGAMBITTS)
    - Learn both reward and treatment model online
  - Partially online GAMBITTS (poGAMBITTS)
    - Learn reward model online, treatment model using offline data

# Simulation Example



# Outline of Paper

- Formalize the GAMBIT framework
- Introduce GAMBITTS
  - Fully and partially online variants
  - Ensemble approaches to support nonlinear modeling
- Theoretical results establishing regret bounds and showing when GAMBITTS has stronger guarantees than standard Thompson sampling
- Simulation experiments, based on the Intern Health Study, investigating the performance of GAMBITTS approaches compared with existing methods