

HELLOPM.

ChatGPT Phase 4 (Assignment 1)

P0 Problem= Discoverability and usage of advanced features (Study Mode / Custom GPTs / Agents)

Goal: Increase activation & engagement by making advanced features discoverable, tryable, and re-entenable for learners and knowledge workers.

P0 Solution: Contextual Micro-Tours — small, intent-triggered demos (3-minute micro-lessons + 1–2 micro-quiz questions) surfaced via banner/spotlight/search/composer to drive first-time activation of Study Mode and other advanced features.

▼ Success metrics

Metric Type	Metric Name	Definition	Why it matters / how it connects to goal
North Star Metric	7-day Feature Activation Rate (feature wise)	% of new users who chose learning as an intent who activate at least one advanced feature (Study Mode, Custom GPT, or Agent) within 7 days of exposure	Best way to know whether discoverability & try flows lead to first-time usage (activation → retention → monetization)
Secondary Metric	Micro-tour Click-through Rate (CTR)	% of users who click "Try Study Mode" banner or spotlight after it's shown	Shows immediate interest from the discovery touchpoint
Secondary	Micro-lesson Completion Rate	% of users who start and complete the micro lesson (mini demo)	Indicates quality & usefulness of the try flow; prerequisite for conversion to full Study Mode
Secondary	Micro-tour → Full Study Mode Conversion	% of micro-tour completers who continue to full Study Mode within same session	Shows the micro tour demo is persuasive or sensible enough to move users deeper into it
Secondary	Avg Session Length (Learner cohort)	Average time spent per session by users with the learning intent	more time = deeper usage (proxy) (+8-10% for the specific cohort)
Secondary	Repeat Feature Usage (14-day re-use)	% of users who reuse an advanced feature within 14 days	Retention for the feature — shows sustained value.
Guardrail	Average API Cost per User	Incremental model call cost per user (USD)	Prevents a pilot from becoming financially unsustainable

Metric Type	Metric Name	Definition	Why it matters / how it connects to goal
Guardrail	User Complaint Rate (discovery / UI confusion)	% of exposed users who submit feedback flagged as negative about discovery UI	Ensures new UI doesn't cause confusion or friction.
Guardrail	Accuracy Flag Rate in Micro-lessons	% of micro lessons flagged for wrong info by QA or user feedback	Safety: micro demos must not propagate incorrect facts
Guardrail	Churn (14 days)	% increase in short-term churn (14 days)	Ensures we're not harming retention overall

Targets (pilot):

- Key Metric: +20% relative increase in 7-day advanced feature activation (treatment vs control)
- Micro-tour CTR $\geq 10\%$ (banner/spotlight)
- Micro-lesson completion $\geq 50\%$ of starters (in the A/B test)
- Micro-tour $\rightarrow$ full conversion ($\geq 25\%$ of completers)
- Shelf CTR $\geq 30\%$ for returning users
- Guardrail: incremental cost per convert (a certain threshold amount) , complaint rate not increased

Pitfalls, assumptions & mitigations

Assumption at Risk	Failure Mode	Impact	Likelihood	Mitigation (fast, non-engineering where possible)	Detection Signal	Containment / Rollback
Users will notice & click the micro-tour banner/spotlight	Low CTR (ignored as "promo")	M	M	A/B test copy & placement; only show after intent signal ("study" keyword); limit frequency	shown $\rightarrow$ clicked CTR < 5%	Pause banner; use inline composer chip only
Micro-lessons are useful & accurate	Micro-lessons contain factual errors / hallucinations	H	M	Conservative generator settings; short outputs; attach source hints; manual QA sample (n=200)	hallucination > 0.5% (or a different decided number) or complaints spike or increased drop-offs/churn	Disable micro-lessons for high-risk topics; fallback to summary + sources
Generation cost is high per exposure	Pilot costs exceed budget	H	M	Cap token size; throttle per user (max 3/week); cache repeat outputs		Reduce token limit; pause micro-lessons; fallback to static summary

Assumption at Risk	Failure Mode	Impact	Likelihood	Mitigation (fast, non-engineering where possible)	Detection Signal	Containment / Rollback
Users find personalization intrusive	High dismiss rate or opt-outs	M	Low - medium	Use simple trigger logic (explicit intent); add "Not for me" dismiss option	Rise in banner dismiss rate	Turn off personalization for the specific cohort & revert to static screen
Micro-tour reduces core flow performance (UI clutter)	Drop in primary chat engagement	M	L	Frequency cap; limit to one surface; show to sampled users only	Drop in chat opens / session abandonment ↑	Reduce exposure %; remove banner from certain pages

Product Requirements Document

Feature: Contextual Micro-Tours (Try Study Mode)

Stakeholders: Design lead, Engineering (Frontend, Backend, QA), Data Analytics, Compliance team (privacy), Growth Manager, Customer Experience/Support head

Goal:

Increase first-time activation of Study Mode among learners by surfacing lightweight, intent-triggered demos that prove value with minimal commitment

▼ Scope of the feature

In scope (for pilot):

- Desktop web banner + search suggestion CTA to open the micro-tour
- Micro-tour modal: setup (level + duration) → micro-lesson generation (2 bullets + 2 MCQs) → success modal (save/remind)
- Rate limit & caching logic, conservative generation config, manual QA sampling
- A/B test for control vs treatment on defined learner cohort

Out of scope (for pilot):

- Full Study Mode product expansion (beyond linking), mobile rollout, ML personalization, GPT library changes, Agent deep integration

▼ User stories

- As a **learner**, when I type "study X", I want a quick demo so I can see structured output and a quiz quickly (under 3 mins)
- As a **returning learner**, I want the micro-tour to save a lesson so I can revisit it later

▼ UX flow

Entry points:

1. Intent detection - when user types / enters "study, quiz, explain, learn"
2. Search suggestion - "Try Study Mode" shown inline for queries with study intent
3. Menu Spotlight - shown/promoted card "Try Study Mode"

Micro-tour modal flow:

- **Modal step 1:** "Start your 3-min demo" – select Level (Beginner / Intermediate / Advanced) and Duration (3 / 10 / 20). Primary CTA: **Start Lesson**
- **Modal step 2 (lesson):** Two-column view – left: 2–3 concise bullets (summary); right: 2 MCQs with radio options + "Check answer" (Progress indicator 1/2)
- **Completion (success):** Card: "Lesson saved to Study Planner." Option: "Send reminder in 3 days" (checkbox). CTAs: **Open Study Planner / Return to chat**

Microcopy examples:

- Banner: "Try Study Mode – 3-minute lessons + quick quizzes" → CTA: "Try Study Mode"
- Modal button: "Start Lesson" / "Save as Flashcards" / "Continue to full Study Mode"

▼ Technical constraints & implementation

- **Generation:** limit tokens (e.g., 150–300 tokens); include a short source sentence when topic is factual
- **Caching:** cache micro-lesson for identical topic/level combos for 'N' hours to reduce cost.
- **Topic gating:** for high-risk topics (medical/legal/financial) fallback to "Short summary + reliable source links" (no quiz) and clearly mention the same (increases trust and reliability)
- **Client:** lightweight modal overlay, accessible keyboard navigation, dismissible.
- **Analytics:** instrument all events with treatment flag and context (topic, intent, placement).

▼ Analytics & events

Instrument the following events and properties:

microtour: shown → clicked → started → answered → completed → expanded_to_full/saved_lesson

Daily dashboard should show risks, CTR, completions, conversion, cost per conversion, cost per generation, complaint counts, and hallucination flags

▼ A/B Test design

- **Population:** users with «learning/study intent» (define using composer keywords + recent behavior)
- Control (no micro-tour) vs Treatment (micro-tour surfaced at detected intent)
- Start 5% (pilot), increase specific cohorts based on signals
- **Duration:** min 2 weeks or until 10k (a certain number) users / statistical threshold reached
- **Primary outcome:** Key Metric = 7-day feature activation rate
- **Secondary outcomes:** micro-lesson completion, micro-tour → full conversion, avg session length
- **Stopping rules:** pause if hallucination flag rate > threshold, cost per conversion > threshold or complaint rate > threshold

▼ Acceptance criteria (for pilot)

- Instrumentation visible in data pipeline and smoke-tested
- Micro-tour CTR ≥ 10% OR micro-lesson completion ≥ 50% (either indicates engagement)

- Micro-tour → full conversion $\geq 25\%$ OR Key Metric uplift statistically significant ($p < 0.05$) depending on sample size
- Incremental cost per convert $\leq$ org threshold; hallucination & complaint guardrails not exceeded
- No critical usability issues from 5 moderated sessions

▼ Monitoring & alerts (first 72 hours)

- Daily dashboard: microtour shown, microtour clicked, micro-lesson completed, feature activated, avg cost per converted user, complaint rate, hallucination flag rate
- Alerts (auto): cost per convert $>$ threshold, hallucination rate $>$ threshold, complaint rate $>$ baseline + 0.5%; auto-notify product & engineering team

▼ Success definition

Pilot considered is successful if:

- Key Metric shows $\geq 20\%$ **relative increase** in 7-day advanced feature activation for treatment vs control & guardrail thresholds not breached.

If not met, we should run 1 iteration (copy/tone/placement changes) and re-test. If it is still negative, we should re-prioritise the solutions and consider alternate discovery solution

▼ Decisions not yet made (***should be decided before before pilot***)

- What is the acceptable generation cost per user? (Finance decision)
 - Which topics are considered high-risk (medical/legal) and should be blocked for micro-lessons or the behaviour in these scenarios should add info like "I might be wrong" with additional source links.
 - Exact cohort definition for "learning/knowledge intent to get the list of intent keywords / signals"
-