# Reddit API Access Request

## Cover Letter - Revised Submission

**Date:** November 25, 2025

**To:** Reddit Data API Team

**From:** Emily Kinney, PhD, Research Staff | Indiana University School of Medicine | emkinney@iu.edu

**Re:** Revised API Access Request (Prior Request #15987717)

## Executive Summary

This is a **revised submission** addressing Request #15987717, which was denied on November 21, 2025 with the reason:

> "Your submission is not in compliance with Reddit's Responsible Builder Policy and/or lacks necessary details."

This cover letter provides comprehensive documentation of our compliance with Reddit's Responsible Builder Policy[16], Developer Terms[14], and Data API Terms[3], with **24 explicit policy citations** demonstrating full alignment with Reddit's requirements.

## Applicant Information

| Field | Value |
| --- | --- |
| **Email** | emkinney@iu.edu[1] |

| | |
|---|---|
| **Role** | Academic Researcher[2] |
| **Reddit Account** | The_MPRINT_Hub |
| **Institution** | Indiana University School of Medicine |
| **Country** | United States |

> ## Detailed Research Proposal
>
> *This section addresses the Reddit form requirement: "Submit a detailed proposal outlining your research purpose, specific data requirements, and ethical considerations."*

# 1. Research Purpose

## Study Overview

This study examines how pregnant individuals discuss medication use, side effects, and healthcare provider interactions on Reddit pregnancy communities. The research is conducted under IRB Protocol #28905 at Indiana University and contributes to NIH-funded research on pregnancy medication safety through the MPRINT Center (NICHD P30HD106451).[2]

## Research Aims

1. **Understand medication information-seeking behavior** in online pregnancy communities
2. **Identify gaps in healthcare communication** that could improve patient education
3. **Analyze peer support patterns** in pregnancy health discussions
4. **Contribute to NIH-funded research** on pregnancy medication safety

## Significance

Pregnant individuals often turn to online communities for medication information due to limited clinical trial data for this population. Understanding these discussions can help healthcare providers better address patient concerns and improve maternal-fetal health outcomes. This research is **non-commercial** and intended solely for peer-reviewed academic publication.[2][12]

## Target Communities

| Subreddit | Subscribers | Research Relevance |
|---|---|---|
| r/pregnant | ~235,000 | Primary pregnancy community |
| r/babybumps | ~350,000 | Large pregnancy support community |
| r/beyondthebump | ~275,000 | Postpartum medication discussions |
| r/tryingforababy | ~130,000 | Pre-conception medication discussions |

## 2. Specific Data Requirements

### Data Elements Collected

| Element | Purpose | Retention |
|---|---|---|
| Post titles | Identify medication discussions | Study duration only |
| Post bodies | Qualitative content analysis | Study duration only |
| Comment text | Understand peer responses | Study duration only |
| Timestamps | Temporal analysis | Study duration only |
| Subreddit name | Community context | Study duration only |
| Vote counts | Community engagement metrics | Study duration only |

All data is accessed via Reddit Data API and processed in compliance with Data API Terms §2.4.[3]

### Data Elements NOT Collected

| Element | Reason |
|---|---|
| Usernames | Immediately anonymized via SHA-256 hash[4] |
| User profiles | Not relevant to research questions |
| Post history | Privacy protection[5] |
| Private messages | Public data only |

| Email addresses | Not accessible, not needed |
|---|---|
| Real names | Not accessible, not needed |
| Location data | Privacy protection |
| Protected Health Information (PHI) | Not collected or inferred |

## Technical Implementation

**API Access:** READ-ONLY via Reddit Data API (OAuth scope: `read` only)[11]

**Rate Limiting:** Token bucket algorithm, 60 requests/minute maximum[17]

**Caching:** LRU cache (100 items) to reduce redundant API calls[18]

**Repository:** https://github.com/stharrold/erkinney-mcp (MIT License, open for audit)[22]

## 3. Ethical Considerations

### IRB Approval

| Field | Value |
|---|---|
| **Protocol Number** | 28905 |
| **Status** | Exempt (Category 4(i) - publicly available information) |
| **Approval Date** | October 31, 2025 |
| **Institution** | Indiana University IRB |
| **Principal Investigator** | Sara Quinney, PhD (squinney@iu.edu) |
| **Research Staff** | Emily Kinney, PhD (emkinney@iu.edu) |

### AoIR Ethics 3.0 Compliance

This research follows the Association of Internet Researchers (AoIR) Ethics Guidelines 3.0:

- **Public data only** - No expectation of privacy in public posts
- **No direct contact** - No interaction with or deception of Reddit users
- **Anonymization** - All user identifiers hashed before analysis[4]

- **Aggregated findings** - No individual quotes without additional ethical review
- **Transparent methodology** - Open-source code for reproducibility

## Privacy-by-Design Implementation

### Zero Tolerance for Privacy Violations

In compliance with Reddit's Responsible Builder Policy[5]:

- SHA-256 username hashing with study-specific salt (irreversible)
- No processing to infer sensitive characteristics (health, political affiliation, etc.)
- No re-identification or de-anonymization attempts
- No matching with off-platform identifiers

## Data Handling and Retention

In compliance with Responsible Builder Policy data handling requirements[6]:

- Data retained only for duration of active research project
- Re-query API for each analysis session (no static database)
- Respect data removals and deletions
- Delete all cached data upon project termination[7]

## No AI/ML Training

### Explicit Compliance Statement

This research tool performs **qualitative content analysis only**. No machine learning models will be trained on Reddit data. The Claude Desktop integration is for **query assistance only** (helping researchers formulate searches), NOT for training AI on Reddit content.

Compliance with:

- Responsible Builder Policy: No Unapproved AI Training[8]
- Developer Terms §4.2[9]
- Data API Terms §2.4 and §3.2[3][10]

### READ-ONLY Commitment

Our tool has **no capability** to:[13]

| Action | Capability |
|---|---|
| Post content | NOT implemented |
| Comment on posts | NOT implemented |
| Vote on content | NOT implemented |
| Send messages | NOT implemented |
| Access private messages | NOT implemented |

## How This Submission Addresses the Prior Denial

| Issue | Resolution |
|---|---|
| **Policy Compliance** | 24 explicit citations to Responsible Builder Policy, Developer Terms, and Data API Terms |
| **Lacking Details** | Comprehensive technical specifications including API endpoints, rate limiting, and anonymization |
| **Submitter Identity** | Submitted by Research Staff (emkinney@iu.edu) with institutional affiliation[1] |
| **Research Context** | Full IRB documentation (Protocol #28905), NIH funding, AoIR Ethics 3.0 compliance |
| **Technical Transparency** | Open-source repository available for audit[22] |

## Policy Compliance Summary

### Privacy Protection

We implement privacy-by-design with SHA-256 username anonymization[4] in compliance with Reddit's zero tolerance for privacy violations.[5] We will never attempt to re-identify, de-anonymize, or reverse engineer data about Redditors.[5]

### Data Handling

Data will be retained only for the duration of the active research project.[6] Upon termination, we will delete all cached or stored User Content.[7] Our tool re-queries the Reddit API for each analysis session rather than maintaining a static database.[6]

### No AI/ML Training

This research tool performs qualitative content analysis only. No machine learning models will be trained on Reddit data, in full compliance with:

- Responsible Builder Policy prohibition on AI training[8]
- Developer Terms §4.2 restrictions[9]
- Data API Terms §2.4 and §3.2 requirements[3][10]

### READ-ONLY Access

Our tool has no capability to post, comment, vote, or send messages.[13] We request only `read` OAuth scope and will access data exclusively through the approved Reddit Data API.[11]

### Rate Limiting

We implement a token bucket rate limiter (60 requests/minute)[17] with LRU caching[18] to minimize API impact, fully respecting Reddit's rate limits.[20]

### Non-Commercial Use

This is federally-funded academic research with no commercial application.[2][12] Research is conducted at a non-profit institution and intended for peer-reviewed publication only.

## Technical Implementation

### Repository

**URL:** https://github.com/stharrold/erkinney-mcp

The complete source code is open-source (MIT License) and available for audit.[22] We will maintain confidentiality of any Reddit confidential information.[23]

### 5 MCP Tools (READ-ONLY)

1. `search_reddit_threads` - Find medication discussions
2. `get_thread_details` - Retrieve thread content for analysis

3. `batch_search_medications` - Efficient multi-medication search
4. `get_subreddit_info` - Document research context
5. `export_research_data` - Export anonymized data for NVivo/Atlas.ti

### Why Not Devvit

This is an offline academic research tool requiring local execution for IRB compliance, not an interactive Reddit application. Devvit does not support this use case.[15]

## Commitments

We commit to:

1. **Transparency** - Full disclosure of purpose, open-source code[16]
2. **Rate Limit Compliance** - 60 req/min token bucket with backoff[10]
3. **No Re-identification** - SHA-256 anonymization, no off-platform matching[5]
4. **Data Deletion** - Delete all data upon project completion[7]
5. **Audit Cooperation** - Full cooperation with any Reddit review[22]
6. **Credential Security** - API credentials stored in macOS Keychain[24]

## Project Status

| Phase | Status |
|---|---|
| Tool Development | Complete |
| IRB Approval | Complete (Oct 31, 2025) |
| API Access Request | This Submission |
| Data Collection | Pending Approval |

**Summary:** Tool complete, IRB approved (Protocol #28905, Oct 2025), NIH-funded (MPRINT P30HD106451); awaiting API approval for 6-month data collection; revised submission addressing Request #15987717.

## Attachments

1. **IRB Approval Letter** - IRB_28905.pdf

2. **Comprehensive Answers Document** - reddit-answers.md
3. **Source Code Repository** - https://github.com/stharrold/erkinney-mcp

## Policy Compliance Matrix

| Requirement | Policy Source | Status |
|---|---|---|
| Approval before access | Responsible Builder Policy[16] | Compliant |
| Transparency | Responsible Builder Policy[16] | Compliant |
| Rate limits | Data API Terms §2.9[20] | Compliant |
| No AI/ML training | RBP[8], Dev Terms[9], Data API[3][10] | Compliant |
| No re-identification | Responsible Builder Policy[5] | Compliant |
| Non-commercial | RBP[2], Dev Terms[12] | Compliant |
| Data retention limits | Responsible Builder Policy[6] | Compliant |
| Query freshness | Responsible Builder Policy[6] | Compliant |
| App review cooperation | Developer Terms §3.1[19] | Compliant |
| Delete on termination | Data API Terms §6[7] | Compliant |
| Bot transparency | Responsible Builder Policy[21] | Compliant |
| Permitted access only | Data API Terms §2.8[11] | Compliant |
| Credential security | Developer Terms §1.4[24] | Compliant |

## Contact Information

| Role | Name | Email |
|---|---|---|
| Research Contact | Emily Kinney, PhD | emkinney@iu.edu |
| Principal Investigator | Sara Quinney, PhD | squinney@iu.edu |

| Developer | Samuel Harrold | samuel.harrold@gmail.com |

Respectfully submitted,

**Emily Kinney, PhD**

Research Staff

Indiana University School of Medicine

MPRINT Center (Maternal & Pediatric Precision in Therapeutics)

emkinney@iu.edu

## Policy Citations (24 Total)

[1] Reddit Researcher API Request Form: *"Please submit this form from a university-affiliated email address or your request will be denied."*

[2] Responsible Builder Policy, "Researchers" section: *"This section applies to academic researchers affiliated with an accredited university and Reddit for Researchers participants who are granted access solely for non-commercial purposes."*

[3] Data API Terms, Section 2.4 "User Content": *"copy and display the User Content using the Data API solely as necessary to develop, deploy, distribute, and run your App"* and *"no other rights or licenses are granted or implied, including any right to use User Content for other purposes, such as for training a machine learning or AI model, without the express permission of rightsholders"*

[4] Implementation: `mcp-bundle-reddit-research/src/privacy/anonymize.js` - SHA-256 username hashing with study-specific salt

[5] Responsible Builder Policy, "Zero Tolerance for Privacy Violations": *"You are strictly prohibited from processing data to derive or infer potentially sensitive characteristics about Reddit users"* and *"you must never attempt to re-identify, de-anonymize, or reverse engineer data about Redditors including by matching data with off-platform identifiers"*

[6] Responsible Builder Policy, "Researchers > Data Handling and Retention": *"Researchers must not retain copies of data beyond what is strictly necessary for the immediate research project"* and *"Researchers must re-run queries against the most recent data export to ensure findings are up-to-date with any data removals or deletions."*

[7] Data API Terms, Section 6 "Termination": *"delete any cached or stored User Content and Materials... This includes any data or models that were derived from User Content"*

[8] Responsible Builder Policy, "No Unapproved Commercialization or AI Training": *"You must not sell, license, share, or otherwise commercialize Reddit data without express written approval. This extends to commercial and non-commercial mining, scraping, or using data for purposes like ads targeting or to train machine learning or AI models."*

[9] Developer Terms, Section 4.2 "Other Use Restrictions": Prohibits using Reddit data *"to train large language, artificial intelligence, or other algorithmic models or related services without our permission"*

**[10]** Data API Terms, Section 3.2 "Restrictions": Prohibits *"circumvent or exceed limitations on calls and use of the Data APIs... or otherwise use the Data APIs in a manner that would constitute excessive or abusive usage"*

**[11]** Data API Terms, Section 2.8 "Permitted Access": *"You will only access (or attempt to access) Data APIs using Access Info described in the Developer Documentation"*

**[12]** Developer Terms, Section 4.1 "Commercial Use Restrictions": Commercial use requires *"express written approval"*

**[13]** Responsible Builder Policy, "Prohibited Bot Activities": *"Bots must not manipulate Reddit's features (e.g., voting, karma) or circumvent safety mechanisms"*

**[14]** Developer Terms, Section 1.3 "Your Information": *"Reddit may ask you to provide us with information about your App... and yourself (e.g., your name and email)... This information must be up to date and accurate at all times."*

**[15]** Responsible Builder Policy, "Developers" section: *"Developers should use the Developer Platform ('Devvit') to build apps on Reddit... If your use case is not supported by Devvit, file a ticket here."*

**[16]** Responsible Builder Policy, "Introduction": *"You must not misrepresent or mask how or why you are accessing Reddit data."*

**[17]** Implementation: `mcp-bundle-reddit-research/src/utils/rate-limiter.js` - Token bucket algorithm, 60 tokens max, 1 token/second refill

**[18]** Implementation: `mcp-bundle-reddit-research/src/utils/cache.js` - LRU cache with 100 item capacity

**[19]** Developer Terms, Section 3.2 "App Users": *"You are solely responsible for your App (including its development, operation, maintenance, support, distribution, use, and App Content)"*

**[20]** Data API Terms, Section 2.9 "API Limitations": *"Reddit may set and enforce limits on your use of the Data APIs"*

**[21]** Responsible Builder Policy, "Bot Transparency": *"Bots should have a clearly specified purpose and scope of access, only accessing the subreddits and API actions they need and which are permitted."*

**[22]** Developer Terms, Section 2.2 "Audit Right": *"Reddit may monitor your use of Developer Services... and audit your App"*

**[23]** Data API Terms, Section 8.1 "Confidentiality": Requirement to maintain confidentiality of Reddit confidential information

**[24]** Developer Terms, Section 1.4 "Access Info": *"You will only access (or attempt to access) and use the Developer Services through... access controls that are authorized and made available to you by Reddit"*

## Policy Document References

**Responsible Builder Policy:** https://support.reddithelp.com/hc/en-us/articles/42728983564564-Responsible-Builder-Policy

**Developer Terms:** https://redditinc.com/policies/developer-terms

**Data API Terms:** https://redditinc.com/policies/data-api-terms